

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:54:18 ; Search time 351.297 Seconds  
(without alignments)  
5302.013 Million cell updates/sec

Title: US-09-990-099-23

Perfect score: 64  
Sequence: 1 caaaaaaagtcgcgcgata.....cccaatttaccatcgtagcc 64

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmbl:  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.pro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vl.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.in.v.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rtd.\*  
36: em.htg.mem.\*  
37: em.htg.vtl.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64	100.0	720	1	AF407018	AF407018 Escherich
2	64	100.0	2886	1	ECOPROT5	M94104 Escherichia
3	64	100.0	10120	1	AE000446	AE000446 Escherich
4	64	100.0	12642	1	AE005600	AE005600 Escherich
5	64	100.0	136254	1	ECOMW82	L10348 E. coli, th
6	64	100.0	221285	2	AC023611	AC023611 Mus muscu
7	64	100.0	252977	2	AC073670	AC073670 Mus muscu
8	64	100.0	256373	2	AC020870	AC020870 Mus muscu
9	64	100.0	272545	2	AC090533	AC090533 Mus muscu
10	64	100.0	280900	1	AP002566	AP002566 Escherich
11	39.6	61.9	245050	1	AL627280	AL627280 Salmonell
12	38	59.4	22492	1	AE008877	AE008877 Homo sapi
13	28.8	45.0	25869	9	AC002459	AC002459 Homo sapi
14	27.4	42.8	66873	2	AC129514	AC129514 Homo sapi
15	27.4	42.8	132288	9	AC099848	AC099848 Homo sapi
16	27.4	42.8	157028	9	AC006442	AC006442 Homo sapi
17	27.4	42.8	157028	9	AL139000	AL139000 Human DNA
18	27.4	42.8	166937	9	AC025679	AC025679 Homo sapi
19	27.4	42.8	178563	2	AP002419	AP002419 Homo sapi
20	27.4	42.8	187559	2	AC119816	AC119816 Mus muscu
21	27.4	42.8	191211	2	AC131053	AC131053 Homo sapi
22	27.4	42.8	317463	2	AC091254	AC091254 Mus muscu
23	27	42.2	98828	9	HSJ388E23	AL049552 Human DNA
24	27	42.2	130110	9	AL731777	AL731777 Human DNA
25	27	42.2	135968	2	AC021707	AC021707 Homo sapi
26	27	42.2	141854	2	AC021595	AC021595 Homo sapi
27	27	42.2	152449	2	AC080175	AC080175 Homo sapi
28	27	42.2	200175	2	AC096004	AC096004 Rattus no
29	26.8	41.9	22516	3	CER50C10	Z72505 Caenorhabdi
30	26.8	41.9	90347	5	AL714030	AL714030 zebrafish
31	26.8	41.9	145634	9	AC096716	AC096716 Homo sapi
32	26.8	41.9	196311	9	CNS01DVP	AL135999 Homo chr
33	26.8	41.9	200598	2	AC012505	AC012505 Homo sapi
34	26.6	41.6	30274	3	CER45H11	Z78420 Caenorhabdi
35	26.6	41.6	82047	3	CER67A10A	AL117206 Caenorhab
36	26.6	41.6	90018	2	AC017853	AC017853 Drosophill
37	26.6	41.6	153439	8	AP003313	AP003313 Oryza sat
38	26.6	41.6	159555	8	AL591710	AL591710 Homo sapi
39	26.6	41.6	162591	9	AL391839	AL391839 Homo sapi
40	26.6	41.6	167026	3	AC010011	AC010011 Drosophill
41	26.6	41.6	182567	2	AC102472	AC102472 Homo muscu
42	26.6	41.6	193902	2	AL591112	AL591112 Homo sapi
43	26.6	41.6	195349	2	AC006705	AC006705 Caenorhab
44	26.6	41.6	195431	2	AC130204	AC130204 Mus muscu
45	26.6	41.6	195879	2	AL807248	AL807248 Mus muscu

## ALIGNMENTS

RESULT 1  
LOCUS AF407018/c 720 bp DNA linear BCT 10-SEP-2001  
DEFINITION Escherichia coli inner membrane protein (ecfI) gene, complete cds.  
ACCESSION AF407018  
VERSION AF407018.1 GI:15529642  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli.  
Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE  
1 (bases 1 to 720)  
AUTHORS Dattigalongue,C., Missiakas,D. and Raina,S.  
TITLE Characterization of the Escherichia coli sigma E regulon  
JOURNAL J. Biol. Chem. 276 (24), 20866-20875 (2001)

MEDLINE 112931005  
PUBMED 11274153  
REFERENCE 2 (bases 1 to 720)  
AUTHORS Dartigalongue, C., Missiakas, D. and Raina, S.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-2001) Biochimie Medicale, CMU, 1 Rue Michel Servet, Geneva 1211, Switzerland  
FEATURES  
source  
1..720  
/organism="Escherichia coli"  
/db\_xref="taxon:562"  
142..>694  
/gene="ecfI"  
142..177  
/gene="ecfI"  
/note="factor sigma24"  
178..>694  
/gene="ecfI"  
/product="inner membrane protein"  
287..694  
/gene="ecfI"  
/note="ecfI; under heat shock transcriptional control of RpoE"  
/codon\_start=1  
/transl\_table=11  
/product="inner membrane protein"  
/protein\_id="AA101385.1"  
/db\_xref="GI:15529643"  
/translation="MMMLFSADRYNSSGQHSRSTDEFIMRNVLAFMIGSMTLGGSSVMSHTGKEGYPGTRASATMIGDETMMGTRSLAIDMPTAVMDLFLPMVDYFRKDSVSRKSEANQATNAVPPARMPDN"  
BASE COUNT 207 a 169 c 183 g 161 t  
ORIGIN  
Query Match 100.0%; Score 64; DB 1; Length 720;  
Best Local Similarity 100.0%; Pred. NO. 6.4e-11;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAAAAAGTCGCTGATTAAGCTTGAAGCTTATTCAGACCCATTTCATCGT 60  
Db 210 CAAAAAAGTCGCTGATTAAGCTTGAAGCTTATTCAGACCCATTTCATCGT 151  
QY 61 AGCC 64  
Db 150 AGCC 147  
RESULT 2  
ECOPROTS 2886 bp DNA linear BCT 26-APR-1993  
LOCUS Escherichia coli ibpA and ibpB genes, complete cds.  
DEFINITION M94104.1 GI:147368  
VERSION ibpA gene; ibpB gene.  
KEYWORDS Escherichia coli (sub-strain W3110, strain K-12) DNA.  
SOURCE Escherichia coli  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 2886)  
AUTHORS Allen, S.P., Polazzi, J.O., Gierse, J.K. and Easton, A.M.  
TITLE Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli  
JOURNAL J. Bacteriol. 174 (21), 6938-6947 (1992)  
MEDLINE 93015757  
PUBMED 1356969  
FEATURES  
source  
location/Qualifiers  
1..2886  
/organism="Escherichia coli"  
/strain="K-12"  
/sub\_strain="W3110"  
/db\_xref="taxon:562"  
539..1083  
/gene="ibpA"

-35\_signal 539..545  
/gene="ibpA"  
/note="putative"  
-10\_signal 560..566  
/gene="ibpA"  
/note="putative"  
RBS 566..660  
/gene="ibpA"  
/note="putative"  
CDS 670..1083  
/gene="ibpA"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA24424.1"  
/db\_xref="GI:147369"  
/translation="MRNFDLSPLRSAGFDRLFNHLENNOSQSGSPPYNLEKSDNHRYRIAIAVAGFAESELITADNDLIVAGAHAEQKERTYLYOGIAERNFERKPOLAENIHVGANLVNGLYIDLEVIPEAKKPRRIEN"  
1181..1623  
/gene="ibpB"  
1181..1186  
/gene="ibpB"  
/note="putative"  
1195..1623  
/gene="ibpB"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA24425.1"  
/db\_xref="GI:147370"  
/translation="MRNFDLSPLRMOWIGFDKLANALONAGESOSFPYNTLEKSDNHRYRIAIAVAGFAESELITADNDLIVAGAHAEQKERTYLYOGIAERNFERKPOLAENIHVGANLVNGLYIDLEVIPEAKKPRRIEN"  
1639..1666  
/note="putative"  
1639..1659  
/note="putative"  
1769..1774  
/note="putative"  
1791..1796  
/note="putative"  
1805..1809  
/note="putative"  
/note="putative"  
1819..2862  
/note="ORF; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA24426.1"  
/db\_xref="GI:147371"  
/translation="MSDIALTVSIIAIVAVVGLTIGNYKFRGIGIGVLFEGIIYVGFVQAGMTSSDMHVIQEFGLIFVYTIQVPGFASLRVSGLELNFAYLIYIGGLVATLHLKLFIDPLPVLVGIIFSGAVTNPALGAGQQLRDGTPEWEDONGMSYAMAPFGICGLIFTMMLRVYERVNVEAOHSSRTNGALIKTINIRVENRMLDLAIDVPTLNGDKTICSRKREPLKVPSPPTIYIQLDLHLVGOPLDHAQVLVIGDEVDSISKTGTDLVERVYVTVNENVLGRIRDLHFKEKYDVVISRLNAGVELVASGDISLORDINAKLL"  
BASE COUNT 693 a 701 c 722 g 770 t  
ORIGIN  
Query Match 100.0%; Score 64; DB 1; Length 2886;  
Best Local Similarity 100.0%; Pred. NO. 5e-11;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAAAAAGTCGCTGATTAAGCTTGAAGCTTATTCAGACCCATTTCATCGT 60  
Db 516 CAAAAAAGTCGCTGATTAAGCTTGAAGCTTATTCAGACCCATTTCATCGT 575  
QY 61 AGCC 64  
Db 576 AGCC 579



```

equal length to YIDS_ECOLI SW: P31456"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/db_xref="GI:1790125"
/translation="MNNVSEKMEHPDVAITIGLPGASALARKLAGKMOYALDKKHQC
GTEGFSKPCGGLADPAORSEFIDGLTLPDVYANPOISVSKVDVAASITRWORSY
ININRAFDLMKSLIPASVEYHDSGLCKRMEDDKMHPEDDQEDHITRIVYG
ADGANSVRLRMHLYPDHDIRKYVAIOQWFKRHKHPERSCIPDSTINCSISKIDGY
FLEGAGYPMKDOTRETTLEKMSAFORPGRTVSEKCTVIFPSWODPVCCKDNAF
LIGGAGFISASISGISTYALDSTDLRSVLLKOPKLTMTAVRATRKRLKLEKIV
KRCULTAPALRKWTRSGVAHIPOLD"
complement(2436..2464)
/feature="factor sigma70: predicted +1 start at 3866982"
complement(2524..2552)
/feature="factor sigma70: predicted +1 start at 3867070"
complement(3512..4849)
/feature="dgot"
/feature="b3691"
complement(3512..4849)
/feature="dgot"
/feature="transport; Transport of small molecules:
carbohydrates, organic acids, alcohols"
/feature="f445: formerly designated yidT"
/codon_start=1
/transl_table=11
/product="D-galactonate transport"
/protein_id="AAC76714.1"
/db_xref="GI:1790126"
/translation="MVGSPAMKIRKIAMOIPVNAKRGRRITLVWIFTVICV
YDRAVLAVASAHIOEERGITKAEKYSASFAMLYTLCIPGWFLDVNGSVYEFIA
IFGMSVATLFGFATGIMSLGLRAITGIFEPAPPTNRWVSWFPEHERASAVGFY
TSGQVGLATPLPLTIQEMLSHWVIVGGIGIIMSLIMFKYOPPLRTKIGSKA
ELDYIRDSGGIVDADPVKKAROPILTAKDKLVKRLIGVLCGFASVSTLMFELT
WPNYLTQEKGITALKAGMTTVPPLAFVGLSGWADLVLRKGFSLGFAKRPIT
CGLISTCIMGANTNDPMIMCLMALAFGNGFASITWSIVSSILAPMLILGTGCVF
NFAGLGGITVPLVGVLAGYGFAPALVYISNAALIALSLITLVGVKRVG"
complement(4924..6687)
/feature="g80a"
/feature="b3692"
complement(4924..6687)
/feature="d90a"
/function="enzyme: Degradation of small molecules: Carbon
compounds"
/feature="f587: formerly designated yidU"
/codon_start=1
/transl_table=11
/product="2-oxo-3-deoxygalactonate 6-phosphate aldolase
and galactonate dehydratase"
/protein_id="AAC76715.1"
/db_xref="GI:2367262"
/translation="MOMOTKLPILAITRGTPPDALAHVCAVITAGDAVEIPLNSPO
WESIPATVAYGKALIGAGTVPKPOVDALARMGCOLVTPNHSVETIRAVGYGM
TVCGCATATBATLERAAGAGENISVIGWSAIGHRSICLRMRDDE
NLMOWIDAGCAGGLSDLYRAGOSVERTAOQAAAFKAVAREAGAMKITKITYRLRP
Query Match 100.0%; Score 64; DB 1; Length 10120;
Best Local Similarity 100.0%; Pred. No. 4e-11;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAAAAAAAGTCGCTGATAAGCGTTGAAAGTCATTTCCAGACCCATTTTACATCGT 60
Db 650 CAAAAAAAAAGTCGCTGATAAGCGTTGAAAGTCATTTCCAGACCCATTTTACATCGT 591
QY 61 AGCC 64
Db 590 AGCC 587

RESULT 4
AE005600/c
LOCUS AE005600 12642 bp DNA linear BCT 21-MAR-2001

```

## DEFINITION

Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 219

of 290.

ACCESSION

AE005600

VERSION

AE005600.1

KEYWORDS

GI:12518521

SOURCE

ORGANISM

Escherichia coli O157:H7 EDL933.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

1 (bases 1 to 12642)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,

Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,

Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,

Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K.,

Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

Welch,R.A. and Blattner,F.R.

Direct Submission

Submitted (22-Oct-2000) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

Location/Qualifiers

1..12642

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/setotype="O157:H7"

/db\_xref="taxon:155864"

/note="enterohemorrhagic"

99..1022

/gene="yidL"

/note="5175"

99..1022

/gene="yidL"

/function="putative regulator; Not classified"

/note="Residues 1 to 307 of 307 are 99.34 pct identical to

residues 1 to 307 of 307 from Escherichia coli K-12 strain

MG1655: B3680"

/codon\_start=1

/transl\_table=11

/product="putative ARAC-type regulatory protein"

/protein\_id="AAG58883.1"

/db\_xref="GI:12518522"

/translation="MSIIRRDIRMNGLQTSVDKRNPPYIPILLINENVSSGISLI

SLMHTYADHERVAVRDRKKPLANSVAVYATVGQKILLKSGEDTILHNGNITFL

KPDNIHSYHSGGLBOYMEFPTGSDMDIPDGGOSYINYEITNOETVEALITSP

EAIKNINAVAFLEKTIITONICIMADGKKDPQRQIKELTATLASLQQRVSVADMA

TIPSEAMLRLEFLITGKTPKEYLDARLDLALSLKQOGNSVGEADVADLPNFSFH

FSKAFKHFEGVAPSAVKNIDHPTDASPHN"

complement(1019..2341)

/gene="25177"

complement(1019..2341)

/gene="25177"

/function="putative enzyme: Transport of small molecules:

carbohydrates, organic acids, alcohols"

/note="Residues 1 to 440 of 440 are 75.05 pct identical to

residues 1 to 441 of 441 from Genpept 118 :

g112145152|gb|AAB63015.1| (U01185) MalH [Fusobacterium

mortiferum]"

/codon\_start=1

/transl\_table=11

/product="putative 6-phospho-beta-glucosidase"

/protein\_id="AAG58884.1"

/db\_xref="GI:12518523"

/translation="MTRFSVYVAGGSGTFPGIVIMLANQDRFPLRAKLFYNDGSR

misc\_feature  
 gene  
 CDS  
 1241..1748  
 /note="O-island #151: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"  
 /complement(2341..3963)  
 /gene="25178"  
 /complement(2341..3963)  
 /gene="25178"  
 /function="putative enzyme: Transport of small molecules: carbohydrates, organic acids, alcohols"  
 /note="Residues 1 to 370 of 540 are 99.45 pct identical to residues 86 to 455 of 455 from Escherichia coli K-12 strain MG1655: B3683"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative PTS component"  
 /protein\_id="AAG58885.1"  
 /db\_xref="GI:12518524"  
 /translation="MEMISQIQRGCGAMFTPVLLPFRAGIVGAILIOMPMFVGSLLTDPNSLFAQIVHIIIEGGWTFERNMPLIFAGLPTGLAKQAQACRACILAVVSPFLSTYFINAGMTWGSYFQGVDFVDVAVAGSLTMAGIKITLDTSTIGAILIGIYTAHLNRLFDKLPVFGIGTSGSYVYIIAFVIMPCAMTLIGMPKVOHGESLCAFLRSAGALVWYVFLERILIPGTHFIEGFIFEGPAVAGGIOMYAOHLQESLAPLSLEPGFALHSGSKTEGANGISLAFYTAAPENRYAGLLIPATLMLVGIETPELPLFISPLFAVHAYLAWSMTVMYIFGVNGMGGILIDVLPDMNIPMSNADMLTQIATGLCTLLYFVFRITLITQNMCTPGKEDAEVYLKSAEYKASRGOTTAAPKEKDQAAGTIOALGAGVGNISINCAATRLRLHMSOTLDEYVFKRLGANGVRSQAIQVIGLHVSQIREQDLSLINSQSAENVATTEAV"  
 4253..4969  
 /gene="yidP"  
 /note="yidP"  
 4253..4969  
 /gene="yidP"  
 /function="putative regulator; Not classified"  
 /note="Residues 1 to 238 of 238 are 99.57 pct identical to residues 1 to 238 of 238 from Escherichia coli K-12 strain MG1655: B3684"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative transcriptional regulator"  
 /protein\_id="AAG58886.1"  
 /db\_xref="GI:12518525"  
 /translation="MIYKSIARLIRLIRLSADFTLNSLLPGSKLAEEFNRSMTIRKAIIDLAMGLVVRHSGTYLVKQVLDLHQTASLITGLVEVLKQKQTVTSOVLIFETMPAPPAISOLARINOEIFSRVREVEKPLMEDSYMPVKLFERNLSLOHLEGSFEYITQPCGILGNGNESLTPVLADELRLARQKVAEHPLRLITSLSYSESEGLNYSVMF RNASEYVEYHRLRHPDKS"  
 complement(4966..6651)  
 /gene="yidP"  
 /note="yidP"  
 /complement(4966..6651)  
 /gene="yidP"  
 /function="putative transport; Not classified"  
 /note="Residues 1 to 561 of 561 are 100.00 pct identical to residues 1 to 561 of 561 from Escherichia coli K-12 strain MG1655: B3685"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative transport protein"  
 /protein\_id="AAG58887.1"  
 /db\_xref="GI:12518526"  
 /translation="MISOEKRTMSDIALTVSIIALAVVGLFGFNKFRGIGIGIGVLEGGIVGHFVSQAGTLDSDMLVIOPEGILLFVYIGIOVGPFPASLRVSGRLNLFVAVLIVIGLVAIALHKLIDPLPVVLGIFSGAVNTPALGAGQIILDLCTPMEMVDOMGMYAMAYPGICGILFTMMMLRVFVNVEAQHSESRKTGGLITITINRVENPLHDLAKDVPICILNGDKITICSRLKRETLTPSPDIIQGDLLHLVGPADHNAQIVIGQVDTSLSKTGDLRVERVVVTVNENYIGKRIIDLHFEKRYDVVISRLNAG

gene  
 CDS  
 1241..1748  
 /note="O-island #151: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"  
 /complement(2341..3963)  
 /gene="25178"  
 /complement(2341..3963)  
 /gene="25178"  
 /function="putative enzyme: Transport of small molecules: carbohydrates, organic acids, alcohols"  
 /note="Residues 1 to 370 of 540 are 99.45 pct identical to residues 86 to 455 of 455 from Escherichia coli K-12 strain MG1655: B3683"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative PTS component"  
 /protein\_id="AAG58885.1"  
 /db\_xref="GI:12518524"  
 /translation="MEMISQIQRGCGAMFTPVLLPFRAGIVGAILIOMPMFVGSLLTDPNSLFAQIVHIIIEGGWTFERNMPLIFAGLPTGLAKQAQACRACILAVVSPFLSTYFINAGMTWGSYFQGVDFVDVAVAGSLTMAGIKITLDTSTIGAILIGIYTAHLNRLFDKLPVFGIGTSGSYVYIIAFVIMPCAMTLIGMPKVOHGESLCAFLRSAGALVWYVFLERILIPGTHFIEGFIFEGPAVAGGIOMYAOHLQESLAPLSLEPGFALHSGSKTEGANGISLAFYTAAPENRYAGLLIPATLMLVGIETPELPLFISPLFAVHAYLAWSMTVMYIFGVNGMGGILIDVLPDMNIPMSNADMLTQIATGLCTLLYFVFRITLITQNMCTPGKEDAEVYLKSAEYKASRGOTTAAPKEKDQAAGTIOALGAGVGNISINCAATRLRLHMSOTLDEYVFKRLGANGVRSQAIQVIGLHVSQIREQDLSLINSQSAENVATTEAV"  
 4253..4969  
 /gene="yidP"  
 /note="yidP"  
 4253..4969  
 /gene="yidP"  
 /function="putative regulator; Not classified"  
 /note="Residues 1 to 238 of 238 are 99.57 pct identical to residues 1 to 238 of 238 from Escherichia coli K-12 strain MG1655: B3684"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative transcriptional regulator"  
 /protein\_id="AAG58886.1"  
 /db\_xref="GI:12518525"  
 /translation="MIYKSIARLIRLIRLSADFTLNSLLPGSKLAEEFNRSMTIRKAIIDLAMGLVVRHSGTYLVKQVLDLHQTASLITGLVEVLKQKQTVTSOVLIFETMPAPPAISOLARINOEIFSRVREVEKPLMEDSYMPVKLFERNLSLOHLEGSFEYITQPCGILGNGNESLTPVLADELRLARQKVAEHPLRLITSLSYSESEGLNYSVMF RNASEYVEYHRLRHPDKS"  
 complement(4966..6651)  
 /gene="yidP"  
 /note="yidP"  
 /complement(4966..6651)  
 /gene="yidP"  
 /function="putative transport; Not classified"  
 /note="Residues 1 to 561 of 561 are 100.00 pct identical to residues 1 to 561 of 561 from Escherichia coli K-12 strain MG1655: B3685"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative transport protein"  
 /protein\_id="AAG58887.1"  
 /db\_xref="GI:12518526"  
 /translation="MISOEKRTMSDIALTVSIIALAVVGLFGFNKFRGIGIGVLEGGIVGHFVSQAGTLDSDMLVIOPEGILLFVYIGIOVGPFPASLRVSGRLNLFVAVLIVIGLVAIALHKLIDPLPVVLGIFSGAVNTPALGAGQIILDLCTPMEMVDOMGMYAMAYPGICGILFTMMMLRVFVNVEAQHSESRKTGGLITITINRVENPLHDLAKDVPICILNGDKITICSRLKRETLTPSPDIIQGDLLHLVGPADHNAQIVIGQVDTSLSKTGDLRVERVVVTVNENYIGKRIIDLHFEKRYDVVISRLNAG

Query Match 100.0%; Score 64; DB 1; Length 12642;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-11;  
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCGCTGATAGGCTTGAAGTTCATTTCCAGACCATTTTACATCGT 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 7930 CAAAAAAGTCGCTGATAGGCTTGAAGTTCATTTCCAGACCATTTTACATCGT 7871  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 ACCC 64  
 ||||  
 DB 7870 AGCC 7867

RESULT 5  
 ECOWM82/c 136254 bp DNA linear BCT 07-FEB-1995  
 LOCUS ECOWM82  
 DEFINITION E. coli; the region from 81.5 to 84.5 minutes.  
 ACCESSION L10328  
 VERSION L10328.1 GI:290484

VELVAGSDLSIORCTILNVGRPSAIDAVANVIGNAQOKLOOVQMLPVYIGIGVLV  
 GSTPVPVPGPALKIGIAGGLIMALLIGRIGISIGKLVGMPFPSPANALRELIVLF  
 LSVVGLKSGDGFVNTLVNEXLSWITGYGLITAVPLITVYIGIARMLAKMNVLTJMGML  
 AGSMTPPALAFANNDPHTSGAALSYATVPLVWFLRIITPOLLAFLVMSIG"  
 complement(6823..7257)  
 /gene="ibpB"  
 /note="ibpB"  
 complement(6823..7257)  
 /gene="ibpB"  
 /function="factor: Adaptations, atypical conditions"  
 /note="Residues 1 to 144 of 144 are 100.00 pct identical to residues 1 to 144 of 144 from Escherichia coli K-12 strain MG1655: B3686"  
 /codon\_start=1  
 /transl\_table=1  
 /product="heat shock protein"  
 /protein\_id="AAG58888.1"  
 /db\_xref="GI:12518527"  
 /translation="MTMKNEDLSPLMRQWIGFDKLANALONAGESQSPPYNTEKSDNHYRTIALAGFROEDLEIOEGRTLSVKGTPQEPQEKRWLHOGILMNQPFSLFTLAENMEVSGATFVNGLIHIDLIRNEPEPIAORIALISERPALNS"  
 complement(7363..7776)  
 /gene="ibpA"  
 /note="ibpA"  
 complement(7363..7776)  
 /gene="ibpA"  
 /function="factor: Adaptations, atypical conditions"  
 /note="Residues 1 to 137 of 137 are 100.00 pct identical to residues 1 to 137 of 137 from Escherichia coli K-12 strain MG1655: B3687"  
 /codon\_start=1  
 /transl\_table=1  
 /product="heat shock protein"  
 /protein\_id="AAG58889.1"  
 /db\_xref="GI:12518528"  
 /translation="MRNFDLSPLYSAGIFDRLFNHLENNOSQNGPYRYNVELVDENHYRTIAVAGFAESELTEITADQNLVYGAHADQKERTYLYOGIARNERFERFOLAENHYVAGLVNGLYIDLERVPEKKPRIEIN"  
 8007..8414  
 /gene="yidQ"  
 /note="yidQ"  
 8007..8414  
 /gene="yidQ"  
 /function="orf: Unknown function"  
 /note="Residues 1 to 135 of 135 are 99.25 pct identical to residues 1 to 135 of 135 from Escherichia coli K-12 strain MG1655: B3688"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AAG58890.1"  
 /db\_xref="GI:12518529"  
 /translation="MMMLFSAARDYNSGQHPSTDEFIMIRNVLLAFNICSQMLLGGCSVMSHTGKGTGTYGTRASVMIIGDETNGTSLAILLMPFTAVAMDLLPLWVD

## KEYWORDS

## SOURCE

Escherichia coli K12 strain MG1655; lambda clones EC14-52, EC17-187, EC27-890, EC21-99, EC22-175, EC27-236, EC17-30, EC17-137, EC19-71, EC15-15, EC17-159, subclones in M13mp19 or Janus.

## ORGANISM

Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

## REFERENCE

1 (bases 1 to 136254)

## AUTHORS

Burland, V., Plunkett, G. III, Daniels, D. L. and Blattner, F. R.

## TITLE

DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication

## JOURNAL

93315143

## MEDLINE

7686882

## COMMENT

This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. A preliminary report was presented at the NIH Human Head meeting/Genome Sequencing and Analysis IV, September 26 - 30, 1992. The entire sequence was independently determined from E. coli MG1655; overlaps and conflicts with other sequence determinations are annotated. The end of this entry overlaps the start of the entry ECOW85 (M87049) by the six bases of an EcoRI site.

Data kindly submitted in computer readable form by: Guy Plunkett III

Laboratory of Genetics  
University of Wisconsin  
445 Henry Mall  
Madison WI 53706  
USA

Phone: 608-262-2534

Email: ecoligenetics.wisc.edu

Fax: 608-263-7459

[1] Authors request hold until publication.

## FEATURES

Location/Qualifiers

1..136254

misc\_feature

/organism="Escherichia coli"

misc\_feature

/db\_xref="taxon:562"

misc\_feature

/note="EC14-52 (lambda clone)"

gene

/complement(<1..669)

CDS

/note="X06036; ECOFP6(1..>669)"

misc\_feature

/gene="fpg"

misc\_feature

/complement(<1..544)

misc\_feature

/gene="fpg"

misc\_feature

/note="alternate gene name mutH"

misc\_feature

/codon\_start=1

misc\_feature

/product="formamidopyrimidine-DNA glycosylase"

misc\_feature

/protein\_id="AA61988.1"

misc\_feature

/db\_xref="GI:643609"

misc\_feature

/translation="MPELPEVETSRGIEPHLVCATILHAVNRGRLEMPVSEIYVSDQVLSVORAKVILLEPGWIIHLSGSGRIIPPELPKPHVDVMSNGVLYRYPDRFGKALMTKEEGHNVLTHGPEPLSDPFGEFTLHQKCAKKAIAIKPWLMDNKLYVGVNIVASLSLAAGI"

misc\_feature

/complement(526..1290)

misc\_feature

/note="J01677; ECORPBG(1..764)"

misc\_feature

/complement(557..586)

misc\_feature

/note="promoter-like sequence; promoter matrix score of 44"

misc\_difference

596..601

misc\_difference

/note="AAAAA in X06036; AAAAA in J01677 and here"

misc\_difference

611

misc\_difference

/note="G in J01677; A in X06036 and here"

misc\_difference

615

misc\_difference

/note="G in J01677; A in X06036 and here"

gene

/note="TTTTTT in X06036; TTTTTT in J01677 and here"

CDS

/complement(642..809)

/gene="rpmG (CG Site No. 236)"

/codon\_start=1

/transl\_table=11

/product="50S ribosomal subunit protein L33"

/protein\_id="AA61989.1"

/db\_xref="GI:290486"

/translation="MAKIREKIKLVSSAGTGHFTTTTKNRPKELEKKFDEVR

OHVIVEKIK"

/complement(830..1066)

/gene="rpmB (CG Site No. 240)"

/complement(830..1066)

/gene="rpmB (CG Site No. 240)"

/codon\_start=1

/transl\_table=11

/product="50S ribosomal subunit protein L28"

/protein\_id="AA61990.1"

/db\_xref="GI:290487"

/translation="MSRCQVTKRRPVTKNNRSHLNATKRRFLPNLHSHRFVSEK

RVTLTVSAKGMVIRKIDTVLAELRAGEKY"

/complement(1283..1957)

/gene="radC (CG Site No. 13913)"

/complement(1283..1957)

/gene="radC (CG Site No. 13913)"

/note="differs from radC sequence in X63366; similar to S.

aureus putative 25 kd protein"

/codon\_start=1

/transl\_table=11

/protein\_id="AA61991.1"

/db\_xref="GI:290488"

/translation="MKYKNSQLMPREKMLKFGISALDVELLALFLRTGTRGKDV

TLAKEMENRSGLYGLTSEYEQSGVHGIGVAKFQALIELARVYVNRREPS

LSPEMTRFLOSLTGEEREIPMVFILDSOHVITHRRLVSGTSLHVVHREITRE

AIKINSALILAHNHPSCAEPKADKLITERIKSCQPMDLRVLDHIYIGCEYVS

AERGWIT"

/complement(1369..1370)

/note="A in X63366; AA here"

/variation

1436

/note="A deleted in radC102 mutation (X63367)"

1698..1822

/note="predicted bend of 76 degrees"

1875..1902

/note="promoter-like sequence; promoter matrix score of 50"

/complement(1911..1940)

/gene="radC (CG Site No. 13913)"

/note="promoter-like sequence; promoter matrix score of 45"

2051..3343

/gene="dfp (CG Site No. 18430)"

2051..3343

/gene="dfp (CG Site No. 18430)"

/codon\_start=1

/transl\_table=11

/protein\_id="AA61992.1"

/db\_xref="GI:290489"

/translation="MKARQKVCXKIANMFWCHPTGIIIMSLAGKIIYVSGGIAAYK

TPELVRLRDRGADRVAVMTAAKAFITPLSLQAVSGPVSDSLDPAEAAGHIDL

GKMDIVLAPRATDLIRVAGAMANDVSTICATPAPVAVLPPANQOAWRAAATOH

NLEVLASRGLIWDGSGOACDGIIGGRKXDDPLTIDMAVAFHSPPNDKIHIMIT

AGPREPLDIPVYISNHSKMGFAIAAARACANTLVSGPISLTPPEVRYDVM

TALEMEAUVNASVOQONIFIGCAVAVYRAATVAPKIKKORQOQDGLITIKMKRNDI

VAGVAALKDRHPYVYVFAETNNVETAROKRIKKNLDLICANDVSQPTGFSNDNA

```

misc_feature
2982..5352
/note="V01578; ECODUTPYR(1..2568)"
misc_feature
2982..4591
/note="X01714; ECOPYREA(1..1609)"
gene
3324..3779
/gene="gut (CG site No. 832)"
3324..3779
/gene="gut (CG site No. 832)"
CDS
3324..3779
/codon_start=1
/transl_table=11
/product="geoxuridinetriphosphatase"
/gene="tkk"
3844..4482
/db_xref="GI:290490"
/gene="tkk"
3844..4482
/function="unknown"
/codon_start=1
/transl_table=11
/product="23 kd protein"
/protein_id="AAA61994.1"
/db_xref="GI:290491"
/translation="MPGKCLSGFVCFMAEKQAKRRREELIQLALTESSDGSQ
RTTAKLASVGSVAALYRHPFSTKRPDSILFIDSLIRINLIKREKDTARL
RITVALLLIGERNFGLRIIRLIGHALMEQDRLGRINOLFERRAQLRVLEKKMR
RLIVLILLIGERNFGLRIIRLIGHALMEQDRLGRINOLFERRAQLRVLEKKMR
EGGVITDTDETLASQTLAFCEGMSLRFVSEFKRPPDDPARPLIAQLQ"
misc_difference
4469..4471
/gene="tkk"
4489..4514
/note="CC in X01714 and V01578; CAC here"
repeat_region
4489..4514
/note="rep (repetitive extragenic palindromic) element;
contains 1 REP sequence"
complement(4519..5160)
/gene="pyre (CG site No. 327)"
complement(4519..5160)
/gene="pyre (CG site No. 327)"
CDS
4469..4471
/codon_start=1
/transl_table=11
/product="rotate phosphoribosyltransferase"
/protein_id="AAA61995.1"
/db_xref="GI:643610"
/translation="MKPYQKQPIEFALSKQVTKFGEFTLKSRKSPYFNGLNTRGR
DLALGFRVAFVALVDGIEFDLFGPAVKGIDVATTTVALAEHHDLDLPYFNKRA
KDHEGNGLVASRGVMLVDVITAGTAIESMEIDIANCATLAVGLISLDQERGR
RGELSAIQEVRDYNCKVISIITLKDLIVYLEEKPEMAHLAAVAYREEEGV"

Query Match
Best Local Similarity 100.0%: Score 64; DB 1; Length 136254;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCGCGATAGGCTTGAAAGTTCATTTCCAGACCATTTTTACATCGT 60
|||||
Db 56970 CAAAAAAGTCGCGATAGGCTTGAAAGTTCATTTCCAGACCATTTTTACATCGT 56911

QY 61 AGCC 64
|||||
Db 56910 AGCC 56907

RESULT 6
AC023611 221285 bp DNA linear HTG 29-JUN-2000
LOCUS AC023611
DEFINITION Mus musculus clone CT7-9K2L, WORKING DRAFT SEQUENCE, 62 unordered
pieces.
AC023611
VERSION AC023611.2 GI:8610277
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Mus musculus.

```

ORGANISM	REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT
----------	---

```

Mus musculus
Mus musculus
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 221285)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 221285)
DOE Joint Genome Institute.
Direct Submission
Submitted (16-PEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 29, 2000 this sequence version replaced gi:6980221.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 956898
Center clone name: RG-MBAC_9K21
-----
Summary Statistics
Consensus quality: 172196 bases at least Q40
Consensus quality: 196011 bases at least Q30
Consensus quality: 201702 bases at least Q20
Estimated insert size: 160000; pulse field gel estimation
Estimated insert size: 215185; sum-of-contigs estimation
Quality coverage: 6.26 in Q20 bases; pulse field gel estimation
Quality coverage: 4.05 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a
* 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1033: contig of 1033 bp in length
1034
1133: gap of unknown length
1134
2542: contig of 1409 bp in length
2543
2642: gap of unknown length
2643
3698: contig of 1056 bp in length
3699
3798: gap of unknown length
3799
4993: contig of 1195 bp in length
4994
5093: gap of unknown length
5094
6427: gap of 1334 bp in length
6428
6527: gap of unknown length
6528
7763: contig of 1236 bp in length
7764
7863: gap of unknown length
7864
8877: contig of 1014 bp in length
8878
8977: gap of unknown length
8979
10256: contig of 1279 bp in length
10257
10356: gap of unknown length
10357
11356: contig of 1000 bp in length
11357
11456: gap of unknown length
11457
12470: contig of 1014 bp in length
12471
13670: gap of unknown length
13671
12571
13861: contig of 1291 bp in length
13862
13961: gap of unknown length
13962
15060: contig of 1099 bp in length
15061
15160: gap of unknown length
15161
16282: contig of 1092 bp in length
16283
16352: gap of unknown length
16353
17470: contig of 1118 bp in length
17471
17570: gap of unknown length
17571
18837: contig of 1267 bp in length
18838
18937: gap of unknown length
20246: contig of 1309 bp in length
20247
21515: gap of unknown length
21516
21515: contig of 1169 bp in length
21516
22752: contig of 1137 bp in length
22752

```

```

22753 22852: gap of unknown length
22853 24438: contig of 1586 bp in length
24439 24538: gap of unknown length
24539 25615: contig of 1077 bp in length
25616 25716: gap of unknown length
25716 26918: contig of 1203 bp in length
26919 27019: gap of unknown length
27019 28161: contig of 1143 bp in length
28162 28262: gap of unknown length
28262 29906: contig of 1645 bp in length
29907 30006: gap of unknown length
30007 31335: contig of 1329 bp in length
31336 31435: gap of unknown length
31436 32661: contig of 1226 bp in length
32662 32761: gap of unknown length
32762 34093: contig of 1332 bp in length
34094 34193: gap of unknown length
34194 35482: contig of 1289 bp in length
35483 35583: gap of unknown length
35583 37002: contig of 1419 bp in length
37002 37101: gap of unknown length
37102 38337: contig of 1236 bp in length
38338 38437: gap of unknown length
38438 39825: contig of 1388 bp in length
39826 39925: gap of unknown length
39926 41162: contig of 1237 bp in length
41163 41262: gap of unknown length
41263 42872: contig of 1610 bp in length
42873 42972: gap of unknown length
42973 44118: contig of 1146 bp in length
44119 44218: gap of unknown length
44219 45913: contig of 1695 bp in length
45914 46013: gap of unknown length
46014 47295: contig of 1282 bp in length
47296 47395: gap of unknown length
47396 47396: contig of 1105 bp in length
47396 48500: gap of unknown length
48501 48600: gap of unknown length
48601 50670: contig of 2070 bp in length
50671 50770: gap of unknown length
50771 52744: contig of 1974 bp in length
52745 52844: gap of unknown length
52845 54280: gap of unknown length
54281 54380: gap of unknown length
54381 56311: contig of 1931 bp in length
56312 56411: gap of unknown length
56412 59166: contig of 2755 bp in length
59167 59266: gap of unknown length
59267 62681: contig of 3415 bp in length
62682 62781: gap of unknown length
62782 65691: contig of 2910 bp in length
65692 65791: gap of unknown length
65792 69358: contig of 3567 bp in length
69359 69458: gap of unknown length
69459 71957: contig of 2499 bp in length
71958 72057: gap of unknown length
72058 76352: contig of 4295 bp in length
76353 76452: gap of unknown length
76453 79737: contig of 3285 bp in length
79738 79837: gap of unknown length
79838 86357: contig of 6520 bp in length
86358 86457: gap of unknown length
86458 90914: contig of 4457 bp in length
90915 91014: gap of unknown length
91015 96832: contig of 5818 bp in length
96833 96932: gap of unknown length
96933 103782: contig of 6850 bp in length
103783 103882: gap of unknown length
103883 110628: contig of 6746 bp in length
110629 110728: gap of unknown length
110729 115501: contig of 5773 bp in length
115502 116501: gap of unknown length
116502 123034: contig of 6433 bp in length
123035 123134: gap of unknown length

```

123135 130209: contig of 7075 bp in length  
130210 130309: gap of unknown length  
130310 141462: contig of 11153 bp in length  
141463 141563: gap of unknown length  
141563 149389: contig of 7827 bp in length  
149389 149489: gap of unknown length  
149489 157307: contig of 7817 bp in length  
157307 157406: gap of unknown length  
157406 167404: contig of 9998 bp in length  
167405 167505: gap of unknown length  
167505 185812: contig of 18308 bp in length  
185813 185912: gap of unknown length  
185913 205026: contig of 19114 bp in length  
205027 205126: gap of unknown length  
205127 221285: contig of 16159 bp in length.

location/Qualifiers  
1. .221285  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="ClbCJ7 mouse BAC library"  
/clone="CT7-9K21"

BASE COUNT 61589 a 46549 c 45795 g 61182 t 6170 others  
ORIGIN  
Query Match 100.0%; Score 64; DB 2; Length 221285;  
Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCGCGTGAATAGGCTTGAAGTTCATTCAGACCATTTACATCGT 60  
DB 42313 CAAAAAAGTCGCGTGAATAGGCTTGAAGTTCATTCAGACCATTTACATCGT 42372  
QY 61 AGCC 64  
DB 42373 AGCC 42376

RESULT 7  
AC073670/c 252977 bp DNA Linear HTG 08-AUG-2000  
LOCUS  
DEFINITION Mus musculus clone CT7-368A6, WORKING DRAFT SEQUENCE, 85 unordered pieces.  
ACCESSION AC073670  
VERSION AC073670.1 GI:8810287  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 252977)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 252977)  
TITLE DOE Joint Genome Institute.  
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 1094499  
Center clone name: RG-MBAC\_368A6  
-----  
Summary Statistics  
Consensus quality: 177673 bases at least Q40  
Consensus quality: 214946 bases at least Q30  
Consensus quality: 224156 bases at least Q20  
Estimated insert size: 100000; pulse field gel estimation  
Estimated insert size: 244577; sum-of-contrigs estimation



Quality coverage: 6.6 in Q20 bases; pulse field gel estimation.  
Quality coverage: 2.7 in Q20 bases; sum of contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 85 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1110: contig of 1110 bp in length  
1111 1210: gap of unknown length  
1211 2457: contig of 1247 bp in length  
2458 2557: gap of unknown length  
2558 3663: contig of 1106 bp in length  
3664 3763: gap of unknown length  
3764 4896: contig of 1133 bp in length  
4897 4997: gap of unknown length  
4998 6152: contig of 1156 bp in length  
6153 6253: gap of unknown length  
6254 7288: contig of 1036 bp in length  
7289 7388: gap of unknown length  
7389 8964: contig of 1576 bp in length  
8965 9064: gap of unknown length  
9065 10174: contig of 1110 bp in length  
10175 10274: gap of unknown length  
10275 11586: contig of 1312 bp in length  
11587 11686: gap of unknown length  
11687 12873: contig of 1187 bp in length  
12874 12973: gap of unknown length  
12974 14204: contig of 1231 bp in length  
14205 14304: gap of unknown length  
14305 15841: contig of 1537 bp in length  
15842 15941: gap of unknown length  
15943 17159: contig of 1218 bp in length  
17160 17259: gap of unknown length  
17260 18555: contig of 1296 bp in length  
18556 18655: gap of unknown length  
18656 19742: contig of 1087 bp in length  
19743 19842: gap of unknown length  
19843 21273: contig of 1431 bp in length  
21274 21373: gap of unknown length  
21374 22750: contig of 1377 bp in length  
22751 22850: gap of unknown length  
22851 24000: contig of 1150 bp in length  
24001 24100: gap of unknown length  
24101 25358: contig of 1438 bp in length  
25359 25638: gap of unknown length  
25639 26726: contig of 1088 bp in length  
26727 26826: gap of unknown length  
26827 28011: contig of 1185 bp in length  
28012 29531: contig of 1420 bp in length  
29532 29631: gap of unknown length  
29632 31152: contig of 1521 bp in length  
31153 31252: gap of unknown length  
31253 32452: contig of 1200 bp in length  
32454 32552: gap of unknown length  
32553 33654: contig of 1102 bp in length  
33655 33754: gap of unknown length  
33755 34774: contig of 1020 bp in length  
34775 34874: gap of unknown length  
34875 36051: contig of 1177 bp in length  
36052 36151: gap of unknown length  
36152 37590: contig of 1439 bp in length  
37591 37690: gap of unknown length  
37691 38908: contig of 1218 bp in length  
38909 39008: gap of unknown length  
39009 40601: contig of 1593 bp in length  
40602 40701: gap of unknown length  
40702 41926: contig of 1225 bp in length  
41927 42026: gap of unknown length  
42028 43983: contig of 1957 bp in length

43984 44083: gap of unknown length  
44084 45206: contig of 1123 bp in length  
45207 45306: gap of unknown length  
45307 46881: contig of 1575 bp in length  
46882 46981: gap of unknown length  
46982 48057: contig of 1075 bp in length  
48057 48157: gap of unknown length  
48157 49669: contig of 1512 bp in length  
49669 51187: gap of unknown length  
51188 51287: gap of unknown length  
51288 52332: contig of 1045 bp in length  
52333 52433: gap of unknown length  
52433 53755: contig of 1322 bp in length  
53755 53855: gap of unknown length  
53855 55582: gap of unknown length  
55582 57262: contig of 1581 bp in length  
57263 57362: gap of unknown length  
57363 59152: contig of 1790 bp in length  
59153 59252: gap of unknown length  
59253 60732: contig of 1480 bp in length  
60733 60832: gap of unknown length  
60833 63030: contig of 2198 bp in length  
63031 63130: gap of unknown length  
63131 65366: contig of 2236 bp in length  
65367 65466: gap of unknown length  
65467 67327: contig of 1761 bp in length  
67328 67327: gap of unknown length  
67328 68506: contig of 1179 bp in length  
68507 68607: gap of unknown length  
68607 71259: contig of 2633 bp in length  
71260 71359: gap of unknown length  
71360 73965: contig of 2606 bp in length  
73966 74065: gap of unknown length  
74066 76125: contig of 2060 bp in length  
76126 76225: gap of unknown length  
76226 77287: contig of 1072 bp in length  
77288 77397: gap of unknown length  
77398 79646: contig of 2249 bp in length  
79647 79746: gap of unknown length  
79747 82449: contig of 2703 bp in length  
82450 82549: gap of unknown length  
82550 85345: contig of 2796 bp in length  
85346 85446: gap of unknown length  
85446 86817: contig of 1372 bp in length  
86817 86917: gap of unknown length  
86918 90145: contig of 3228 bp in length  
90146 90245: gap of unknown length  
90246 91490: contig of 1243 bp in length  
91491 91590: gap of unknown length  
91591 94935: contig of 3345 bp in length  
94936 95035: gap of unknown length  
95036 97845: contig of 2810 bp in length  
97846 97945: gap of unknown length  
97946 100449: contig of 2504 bp in length  
100450 100549: gap of unknown length  
100550 103583: contig of 3034 bp in length  
103584 103683: gap of unknown length  
103684 107678: contig of 3995 bp in length  
107679 107778: gap of unknown length  
107779 110534: contig of 2756 bp in length  
110535 110634: gap of unknown length  
110635 113491: contig of 2857 bp in length  
113492 113591: gap of unknown length  
113592 115603: contig of 2012 bp in length  
115604 115703: gap of unknown length  
115704 118957: contig of 3254 bp in length  
118958 119057: gap of unknown length  
119059 124203: contig of 5146 bp in length  
124204 124303: gap of unknown length  
124304 127811: contig of 3508 bp in length  
127812 127911: gap of unknown length

\* 127912 132157: contig of 4246 bp in length  
\* 132158 132257: gap of unknown length  
\* 132258 135753: contig of 3496 bp in length  
\* 135754 135853: gap of unknown length  
\* 135854 140373: contig of 4520 bp in length  
\* 140374 140473: gap of unknown length  
\* 140474 145161: contig of 4688 bp in length  
\* 145162 145261: gap of unknown length  
\* 145262 149776: contig of 4515 bp in length  
\* 149777 149876: gap of unknown length  
\* 149877 156082: contig of 6206 bp in length  
\* 156083 156183: gap of unknown length  
\* 156183 163129: contig of 6947 bp in length  
\* 163129 163229: gap of unknown length  
\* 163230 167273: contig of 4044 bp in length  
\* 167274 167373: gap of unknown length  
\* 167374 175126: contig of 7753 bp in length  
\* 175127 175226: gap of unknown length  
\* 175227 182114: contig of 6888 bp in length  
\* 182115 182214: gap of unknown length

Query Match  
Best Local Similarity 100.0%; Score 64; DB 2; Length 252977;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCGCTGATAGCGTTGAAAGTCATTCCAGACCATTTTTCATCGT 60  
Db 11123 CAAAAAAGTCGCTGATAGCGTTGAAAGTCATTCCAGACCATTTTTCATCGT 11064

QY 61 AGCC 64  
Db 11063 AGCC 11060

RESULT 8  
LOCUS AC020870/c 256373 bp DNA linear HTG 17-FEB-2000  
DEFINITION Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC020870  
VERSION AC020870.2 GI:6984372  
KEYWORDS HTG; HTGS.PHASE0.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 256373)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 256373)  
REFERENCE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
TITLE Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL On Feb 17, 2000 this sequence version replaced g1:6868438.  
COMMENT \* NOTE: This record contains 198 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 752: contig of 752 bp in length  
\* gap of unknown length  
\* 753 1000: contig of 248 bp in length  
\* gap of unknown length  
\* 1001 1604: contig of 604 bp in length  
\* gap of unknown length  
\* 1605 2302: contig of 698 bp in length

\* 2303 2699: contig of 397 bp in length  
\* gap of unknown length  
\* 2700 3686: contig of 987 bp in length  
\* gap of unknown length  
\* 3687 4099: contig of 413 bp in length  
\* gap of unknown length  
\* 4100 4778: contig of 679 bp in length  
\* gap of unknown length  
\* 4779 5571: contig of 793 bp in length  
\* gap of unknown length  
\* 5572 5967: contig of 396 bp in length  
\* gap of unknown length  
\* 5968 6763: contig of 796 bp in length  
\* gap of unknown length  
\* 6764 7639: contig of 876 bp in length  
\* gap of unknown length  
\* 7640 8442: contig of 803 bp in length  
\* gap of unknown length  
\* 8443 9414: contig of 972 bp in length  
\* gap of unknown length  
\* 9415 10236: contig of 822 bp in length  
\* gap of unknown length  
\* 10237 10844: contig of 608 bp in length  
\* gap of unknown length  
\* 10845 11504: contig of 660 bp in length  
\* gap of unknown length  
\* 11505 12321: contig of 817 bp in length  
\* gap of unknown length  
\* 12322 12652: contig of 331 bp in length  
\* gap of unknown length  
\* 12653 14116: contig of 1464 bp in length  
\* gap of unknown length  
\* 14117 15298: contig of 1182 bp in length  
\* gap of unknown length  
\* 15299 16234: contig of 936 bp in length  
\* gap of unknown length  
\* 16235 16376: contig of 142 bp in length  
\* gap of unknown length  
\* 16377 16664: contig of 288 bp in length  
\* gap of unknown length  
\* 16665 17631: contig of 967 bp in length  
\* gap of unknown length  
\* 17632 19241: contig of 1610 bp in length  
\* gap of unknown length  
\* 19242 20453: contig of 1212 bp in length  
\* gap of unknown length  
\* 20454 20559: contig of 106 bp in length  
\* gap of unknown length  
\* 20560 21117: contig of 558 bp in length  
\* gap of unknown length  
\* 21118 21655: contig of 538 bp in length  
\* gap of unknown length  
\* 21656 21818: contig of 163 bp in length  
\* gap of unknown length  
\* 21819 22679: contig of 861 bp in length  
\* gap of unknown length  
\* 22680 23579: contig of 900 bp in length  
\* gap of unknown length  
\* 23580 24383: contig of 804 bp in length  
\* gap of unknown length  
\* 24384 25612: contig of 1229 bp in length  
\* gap of unknown length  
\* 25613 25957: contig of 345 bp in length  
\* gap of unknown length  
\* 25958 27260: contig of 1303 bp in length  
\* gap of unknown length  
\* 27261 28596: contig of 1336 bp in length  
\* gap of unknown length  
\* 28597 29473: contig of 877 bp in length  
\* gap of unknown length  
\* 29474 30372: contig of 899 bp in length  
\* gap of unknown length

```

* 30373 31202: contig of 830 bp in length
* 31203 32688: contig of 1486 bp in length
* 32689 33431: contig of 743 bp in length
* 33432 33693: contig of 262 bp in length
* 33694 34616: contig of 923 bp in length
* 34617 35884: contig of 1268 bp in length
* 35885 36459: contig of 575 bp in length
* 36460 36996: contig of 537 bp in length
* 36997 37941: contig of 945 bp in length
* 37942 38682: contig of 741 bp in length
* 38683 39549: contig of 867 bp in length
* 39550 40792: contig of 1243 bp in length
* 40793 42585: contig of 1793 bp in length
* 42586 44115: contig of 1530 bp in length
* 44116 45300: contig of 1185 bp in length
* 45301 46417: contig of 1117 bp in length
* 46418 47815: contig of 1398 bp in length
* 47816 49310: contig of 1495 bp in length
* 49311 50304: contig of 994 bp in length
* 50305 50848: contig of 544 bp in length
* 50849 51828: contig of 980 bp in length
* 51829 52699: contig of 871 bp in length
* 52700 53365: contig of 666 bp in length
* 53366 54633: contig of 1268 bp in length
* 54634 55470: contig of 837 bp in length
* 55471 56358: contig of 888 bp in length
* 56359 56934: contig of 576 bp in length
* 56935 57672: contig of 738 bp in length
* 57673 58319: contig of 647 bp in length
* 58320 58529: contig of 210 bp in length
* 58530 59083: contig of 554 bp in length
* 59084 59906: contig of 823 bp in length
* 59907 60826: contig of 920 bp in length
* 60827 61084: contig of 258 bp in length
* 61085 61762: contig of 678 bp in length
* 61763 62638: contig of 876 bp in length
* 62639 62713: contig of 75 bp in length

```

```

Query Match      100.0%; Score 64; DB 2; Length 256373;
Best Local Similarity 100.0%; Pred. No. 2,3e-11;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 CAAAAAAGTCCGCTGATTAAGCTTGAAGTTTCATTTCCAGACCCATTTTTACATCGT 60
Db 31717 CAAAAAAGTCCGCTGATTAAGCTTGAAGTTTCATTTCCAGACCCATTTTTACATCGT 31658
OY 61 AGCC 64
Db 31657 AGCC 31654

```

```

RESULT 9
AC090533 272545 bp DNA linear HTG 15-MAY-2002
LOCUS Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT
DEFINITION AC090533.4 GI:17976439
AC090533
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.

```

```

REFERENCE 1 (bases 1 to 272545)
AUTHORS Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
High Throughput Mouse Sequencing

```

```

TITLE Unpublished
JOURNAL 2 (bases 1 to 272545)
REFERENCE Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,
AUTHORS Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
Direct Submission
Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Dec 21, 2001 this sequence version replaced gi:13310871.

```

## COMMENT

```

-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcg.org/Sequence/mouse.html
Contact: hpcg@meda1.mgh.harvard.edu
-----Summary Statistics
Center project name: ADU
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 257441 at least 020
*Consensus quality: 253718 at least 030
*Consensus quality: 247095 at least 040
*Estimated insert size: agarose-FP - N/A
**Estimated insert size: 271655 - sum-of-coverage
Quality coverage: agarose-FP - N/A

```

\* This is a working draft sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence,  
\* as soon as it is available and the accession number will  
\* be preserved.

[illegible]

FEATURES	source
*	260460 260479: gap of unknown length
*	260480 260829: contig of 350 bp in length
*	260830 260849: gap of unknown length
*	260850 261661: contig of 812 bp in length
*	261662 261681: gap of unknown length
*	261682 262532: contig of 851 bp in length
*	262533 262552: gap of unknown length
*	262553 263351: contig of 799 bp in length
*	263352 263371: gap of unknown length
*	263372 264217: contig of 846 bp in length
*	264218 264237: gap of unknown length
*	264238 265205: contig of 968 bp in length
*	265206 265225: gap of unknown length
*	265226 266270: contig of 1045 bp in length
*	266271 266290: gap of unknown length
*	266291 266749: contig of 459 bp in length
*	266750 266769: gap of unknown length
*	266770 267718: contig of 949 bp in length
*	267719 268579: gap of unknown length
*	268577 268596: gap of 838 bp in length
*	268597 270010: contig of 1414 bp in length
*	270011 270030: gap of unknown length
*	270031 271654: contig of 1624 bp in length
*	271655 271674: gap of unknown length
*	271675 272545: contig of 871 bp in length.
	location/Qualifiers
	1..272545

```
misc_feature 1. 30267
              /note="assembly_name:Contig108
misc_feature 30288. 58400
              /note="assembly_name:Contig107
              clone_end:SP6
              vector_side:left"
misc_feature 58421. 88864
              /note="assembly_name:Contig106
              88885. 109889
              /note="assembly_name:Contig105"
              109910. 125133
              /note="assembly_name:Contig104
              clone_end:77
              vector_side:right"
misc_feature 125154. 139549
              /note="assembly_name:Contig103"
              139570. 153410
              /note="assembly_name:Contig102"
              153431. 162857
              /note="assembly_name:Contig101"
              162888. 171044
              /note="assembly_name:Contig100"
              171065. 179304
              /note="assembly_name:Contig99"
              179325. 187059
              /note="assembly_name:Contig98"
              187080. 197598
              /note="assembly_name:Contig97"
              197619. 205966
              /note="assembly_name:Contig96"
              205987. 212449
              /note="assembly_name:Contig95"
              212470. 217752
              /note="assembly_name:Contig94"
              217773. 223943
              /note="assembly_name:Contig93"
              223964. 230362
              /note="assembly_name:Contig92"
              230383. 236251
              /note="assembly_name:Contig91"
```





```

/note="Orthologue of E. coli glnA (GlnA_ECOLI): Fasta hit
to GlnA_ECOLI (468 aa), 98% identity in 468 aa overlap"
/codon_start=1
/transl_table=11
/product="glutamine synthetase"
/protein_id="CAD03093.1"
/db_xref="GI:16504730"
/db_xref="SWISS-PROT:P06201"
/translation="MSAEVLTMLNEHEVKKVVDLRFDTDKGEKHVHTIPAHQVAEEF
EGKMFDSIGGKMGKINSDVLMPEADSTAVIDFEDSTLLICDILEGTCQGD
RDPRIAGKAEYLRATGADIVLFGPEEFELPDIDREFGASISGSHVADIDIGAN
SSTKYEGKKGKRGKGGVFPVPSAODIRSEMCLVMEOMGLVFAHHEVATG
OMEVATREMTKKADEIOIKYVYVNAHREKATPMKPMFGDNGSGMCHMSLA
KSTNIFSDCKAGLSEOLYIGVIVKAKINLANLPNTSYKRLVGYEAVVMA
YSNRKSASIRIPVVASPKARIRIIVRPDPANPYLCFALLMAGLDCKIKIHGEA
MDKNLYDLPEEKELPOVAGSLEALNALDIDREFLAKAGVFIDEADIVAIAREE
MDRVMTPHPPEFELYYSV"
/gene="STY3874"
/note="Pfam match to entry PF00120, gln-synt, Glutamine
synthetase, score 838.00, E-value 2.4e-266"
/region_start=1
/region_end=522
/gene="STY3874"
/note="PS00180 Glutamine synthetase signature 1"
/gene="STY3874"
/note="PS00181 Glutamine synthetase putative ATP-binding
region signature"
/region_start=1474
/region_end=1512
/gene="STY3874"
/note="PS00182 Glutamine synthetase class-1 adenylation
site"
/gene="STY3874"
/note="STY3875"
/region_start=2003
/region_end=3052
/gene="STY3875"
/note="STY3875"
/region_start=2003
/region_end=3052
/gene="STY3875"
/note="Orthologue of E. coli glnL (NTRB_ECOLI): Fasta hit
to NTRB_ECOLI (349 aa), 93% identity in 349 aa overlap"
/codon_start=1
/transl_table=11
/product="two-component system sensory histidine kinase"
/protein_id="CAD03094.1"
/db_xref="GI:16504731"
/db_xref="SWISS-PROT:P41788"
/translation="MASGIOPDAGQIINSIVLVVDDALAIHYANPAOQLAQS
RLFTGPIPELLSTFSLNIDIMRESLAGSGFTNNEVTIVDSRSHLSIAORLPD
FILEMAPMDNORLSQEQLOHQAQIDARLDVRLGAIHETKIPVGLGSAOILSKALP
DPALEYKVIIEQADRLNLVDRILGPHGMIHETIRHVARVYALVSMELPNV
RLIRYDPSLEPLPHDPEQIEQVILNIRALQALGPGCGEITLRTLRDQLHGER
VRLARIVEDNGPGIPHLQDTLEYPWVSGREGTGIGLSIARNLLIDHAGKLEFIS
WPGHTEFSVYPIRK"
/region_start=2417
/region_end=3040
/gene="STY3875"
/note="Pfam match to entry PF00512 signal, Histidine
kinase, score 244.40, E-value 1.6e-69"
/region_start=3061
/region_end=4470
/gene="STY3876"
/note="glnG"
/region_start=3061
/region_end=4470
/gene="STY3876"
/note="STY3876"
/region_start=3061
/region_end=4470
/note="Fasta hit to HYDG_ECOLI (441 aa), 44% identity in
462 aa overlap"
/note="Fasta hit to YGNA_ECOLI (529 aa), 36% identity in 451 aa
overlap"
/note="Fasta hit to ATOC_ECOLI (461 aa), 43% identity in 468 aa
overlap"
/note="Fasta hit to YFHA_ECOLI (444 aa), 44% identity in 381 aa
overlap"
/note="Orthologue of E. coli glnG (NTRC_ECOLI): Fasta hit to
NTRC_ECOLI (469 aa), 96% identity in 469 aa overlap"
/codon_start=1
/transl_table=11

```

---

```

/product="Two-component system, response regulator"
/protein_id="CAD03095.1"
/db_xref="GI:16504732"
/db_xref="SWISS-PROT:P08263"
/translation="MQRGIIVVDDSSIRWVLERALAGLTCITEENGNEVIALA
SKTPVILSDIIMPQMGDIALKQKORHPMLPIVIMAHSDADAAVSAVOQGFEDL
PRPFIDENVAIVARAISHYOEOOPRIENEVNGPTDMIGAPAMOVFRILIGKLS
SISVLINGSGTGKELVNAHRLHRSPRKAPFALNMAALPKDILESLEFGHGAT
GANTTIRGREGQADGTLFDLIDGMPDLVDVOTRILRYADCOFYRGVGLVAVRGI
IAATHONLERRVQEKFRFEDLFRNLVRIHPLPERRERDIPRARHFLQVAREIG
VEAKLHPETELATRLAMPGNVROLNENTCRMLVMAQGVETODDPELFASAD
SPSHLPDPMATLILAMPADRALRSCHONLSEAPBELRTLLTALRHTQHKQBAK
LIGMGNTLIRLKLKELGME"
/region_start=3070
/region_end=3405
/gene="STY3876"
/note="Pfam match to entry PF00072 response_reg, Response
regulator receiver domain, score 141.50, E-value 1.5e-38"
/region_start=3478
/region_end=4173
/gene="STY3876"
/note="Pfam match to entry PF00158 sigma54, Sigma-54
transcription factors, score 515.80, E-value 3.1e-151"
/region_start=3550
/region_end=3591
/gene="STY3876"
/note="PS00675 Sigma-54 interaction domain ATP-binding
region A signature"
/region_start=3736
/region_end=3783
/gene="STY3876"
/note="PS00676 Sigma-54 interaction domain ATP-binding
region B signature"
/region_start=4117
/region_end=4146
/gene="STY3876"
/note="PS00688 Sigma-54 interaction domain C-terminal part
signature"
/complement(4783..6156)
/gene="STY3877"
/note="STY3877"
/note="hemN"
/complement(4783..6156)
/gene="STY3877"
/note="Orthologue of E. coli hemN (HEMN_ECOLI): Fasta hit
to HEMN_ECOLI (457 aa), 93% identity in 457 aa overlap"
/codon_start=1
/transl_table=11
/product="oxygen-independent coproporphyrinogen III
oxidase"
/protein_id="CAD03096.1"
/db_xref="GI:16504733"
/db_xref="SWISS-PROT:P37129"
/translation="WSEQIDMDLALIKYKSGPRYSPTALESEDEPDAFLQA
VARPERPILSVVLPFCHKICFCGCKIVTFQRKADQDLDALEQIIRHAPLPAD
RHVSQILHMGCGPTVILNKAOISRLMTLRENFHFWTDAEISIEVPREIELVDIHR
AGEFNLSMGVDFNKEVORLVNREODEFEFALNHAARDIGFTSTNIDLIGLKOT
PESFAFTLRKRVTELNDRLSYVNAHLPTLPAQKRIKDALPSAQKLLDLOETIVS
LNDAGOPTKRGMDHFAFPDDETLAAVAREGLHREFGYTGQDPLDLCGVSASIMID
GIWONKELKRYOQVDERGNALMRGILTTRDPCIRRVIVIKALICNFDNNAEOMW
GLHFACTYAEADQLQLSLAKDGLVDSKGIQVYAKGLLLRNICMCFDAYLRQAKM
QOFSRVI"
/complement(6345..6860)
/gene="STY3878"
/note="STY3878"
/note="yihI"
/complement(6345..6860)
/gene="STY3878"
/note="Orthologue of E. coli yihI (YIH_ECOLI): Fasta hit
to YIH_ECOLI (169 aa), 85% identity in 169 aa overlap.
Note this protein is extremely hydrophilic"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD03097.1"
/db_xref="GI:16504734"
/db_xref="SWISS-PROT:P37130"
/translation="MKRPTSPAPSKAFGKORRKTRELNQEARDRKRLKHHGHAGS
RAAGNSAGSGGNOQKQDPRISKTPVPLGVETQOQHKPRSEKPMHSPAEIDL
ETDERDALLERLEAGETLSAEDQAVDAKLRIDELKQKLGUSTYDDDEDEDEBQK

```



gene  
EDMRLRLGNGN  
complement(6979, .7203)  
/gene="STV3879"  
complement(6979, .7203)  
/gene="STV3879"  
/note="no significant database hits. Note this protein is extremely hydrophilic"  
/codon\_start=1

Query Match 61.9%; Score 39.6; DB 1; Length 245050;  
Best Local Similarity 77.4%; Pred. No. 0.0031;  
Matches 48; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCCGCTGATAGCGCTGAAAGTTCACATTCACAGCCATTTTACATCGT 60  
Db 110239 CAAAAAAGTCCGCTGATAGCGCTGAAAGTTCACATTCACAGCCATTTTACAGCGT 110298  
QY 61 AG 62  
Db 110239 AG 110300

RESULT 12  
AE008877/c 22492 bp DNA linear BCT 31-JUL-2002  
LOCUS Salmoneella typhimurium LT2, section 181 of 220 of the complete  
DEFINITION genome.  
ACCESSION AE008877 AE006468  
VERSION AE008877.1 GI:16422367  
KEYWORDS  
SOURCE Salmoneella typhimurium LT2.  
ORGANISM Salmoneella typhimurium LT2.  
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Salmoneella.

REFERENCE 1 (bases 1 to 22492)  
AUTHORS McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,  
Lateille, P., Courtney, U., Porwollik, S., Ali, J., Dante, M., Du, F.,  
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,  
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,  
Stonking, T., Nhan, M., Waterston, R. and Wilson, R.K.  
TITLE Complete genome sequence of Salmoneella enterica serovar Typhimurium  
LT2  
JOURNAL Nature 413 (6858), 852-856 (2001)  
MEDLINE 21534948  
PUBMED 11677609  
REFERENCE 2 (bases 1 to 22492)  
AUTHORS The Salmoneella typhimurium Genome Sequencing Project.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of  
Genetics, Washington University School of Medicine, 4444 Forest  
Park Boulevard, St. Louis, MO 63108, USA  
COMMENT Supported by NIH grant 50 01 A143283

Coding sequences below are predicted from manually evaluated  
computer analysis, using similarity information and the programs:  
GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and  
GeneMark: <http://opal.biology.gatech.edu/geneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto  
Encyclopedia of Genes and Genomes: <http://www.genome.ad.jp/kegg/>,  
and Pedro Romero and Peter Karp at EcoCyc:  
<http://ecocyc.org/PangoSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites  
were kindly provided by Heladia Saigado, Julio Collado-Vides and  
Regunob: [http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence

FEATURES  
source from more than one m13 subclone.  
Location/Qualifiers

1..22492  
/organism="Salmoneella typhimurium LT2"  
/strain="LT2; SCSG 1412; ATCC 700720"  
/db\_xref="ATCC:700720"  
/db\_xref="taxon:99287"  
/note="LT2"  
complement(78, .278)  
/gene="tyvL"  
/note="STM3797"  
complement(78, .176)  
/gene="tyvL"  
/note="similar to E. coli tyvB operon leader peptide  
(AAC76695.1); Blastp hit to AAC76695.1 (32 aa), 84%

complement(181, .186)  
/gene="tyvL"  
/note="putative RBS for tyvL; RegulonDB: STMS1H003766"  
complement(254, .262)  
/gene="tyvL"  
/note="putative -10 signal for tyvL;  
RegulonDB: STMLTH004613"  
complement(264, .292)  
/note="putative binding site for CRP, RegulonDB:  
STMS1H000038"  
/bound\_moiety="CRP"  
complement(270, .278)  
/gene="tyvL"  
/note="putative -35 signal for tyvL;  
RegulonDB: STMLTH004613"  
699, .813  
/gene="tyvSdA"  
/note="STM3797A"  
join(699, .790, 789, .813)  
/gene="tyvSdA"  
/note="ribosomal slippage"  
/codon\_start=1  
/transl\_table=11  
/product="tyvSdA"  
/protein\_id="AAM89474.1"  
/db\_xref="GI:22023623"  
/translation="MSTRFLAKGIACATAGNROKGSAGSAPLNHRRV"  
810, .899  
/gene="tyvSdB"  
/note="STM3796B"  
810, .899  
/gene="tyvSdB"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein in the LexA regulon"  
/protein\_id="AAL22656.1"  
/db\_xref="GI:16422369"  
/translation="MSVVDITILIKLVAALQILDVAVLK"  
992, .1852  
/gene="STM3796A"  
992, .1852  
/gene="STM3796A"  
/codon\_start=1  
/transl\_table=11  
/product="integral membrane protein"  
/protein\_id="AAL22657.1"  
/db\_xref="GI:16422370"  
/translation="MTPSTPDANTLSVFCILFAALLHSAWNAIVAGNDKLYAIG  
VSGSAVALILPFSQPAHASICPLAASALAOVYTVVAKTVOVSDMSQVPIAMR  
GTAPLVALISVLEIGDSLISLAWVGIAVCMATIGACGRRASRGCVLATATNCF  
IAGTVLDGTVGRLESTALGVTLSFPLNACLLTMMATARRRARSRYLAQWKKGIF





between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTB-18805 is from the first release of the human BAC library  
CTB-978SK-B. The library contains cloned DNA from the male  
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad.  
Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8  
(1996). This clone is available from Research Genetics, Inc.  
(<http://www.resgen.com>).  
VECTOR: pBelOBAC11  
Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTA-313E3, 200 bp overlap.  
Actual start of this clone is at base position 122698 of CTA-313E3;  
actual end is at base position 25869 of CTB-18805.

#### FEATURES

SOURCE  
1. 25869  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7g22-q31"  
/clone\_id="CTB-18805"  
/clone\_lib="CTB-978SK-B"  
2. 466  
/rpt\_family="L1"  
459..520  
/rpt\_family="L1"  
521..826  
/rpt\_family="Alu"  
803..826  
/rpt\_family="(A)n"  
827..1564  
/rpt\_family="L1"  
1622..1960  
/rpt\_family="AchoBo"  
1961..2274  
/rpt\_family="ERV1"  
2431..2548  
/rpt\_family="AchoBo"  
2442..2469  
/rpt\_family="AT-rich"  
2549..2870  
/rpt\_family="L1"  
2871..2936  
/rpt\_family="AchoBo"  
3128..3424  
/rpt\_family="Alu"  
4043..4236  
/rpt\_family="MIR"  
4209..4239  
/rpt\_family="AT-rich"  
6203..6523  
/rpt\_family="Alu"  
6322..6350

misc\_feature  
6709..6824  
/note="match to EST BF592945 (NTD:911685269)"  
repeat\_region  
7093..7156  
/rpt\_family="L1"  
8790..9147  
/rpt\_family="MALR"  
10502..10744  
/rpt\_family="MALR"  
11131..11158  
/rpt\_family="AT-rich"  
11322..11481  
/rpt\_family="MIR"  
13163..13191  
/rpt\_family="(TTTA)n"  
13279..13543  
/rpt\_family="L2"  
13629..13771  
/rpt\_family="L1"  
13829..14233  
/rpt\_family="L1"  
14234..14767  
/rpt\_family="L1"  
14691..14714  
/rpt\_family="AT-rich"  
14772..16354  
/rpt\_family="L1"  
16355..16869  
/rpt\_family="L1"  
16889..16974  
/rpt\_family="L1"  
16983..17258  
/rpt\_family="Alu"  
17382..17407  
/rpt\_family="AT-rich"  
17582..17731  
/rpt\_family="L2"  
18560..18705  
/rpt\_family="L1"  
20341..20633  
/note="match to EST BG773033 (MID:914083686)"  
20539..20745  
/rpt\_family="MIR"  
21127..21603  
/rpt\_family="ERV1"  
21608..21914  
/rpt\_family="Alu"  
21917..22190  
/rpt\_family="L1"  
22189..22218  
/rpt\_family="(TTTG)n"  
22200..22371  
/rpt\_family="Alu"  
22421..22977  
/rpt\_family="ERV1"  
23267..23429  
/rpt\_family="Alu"  
23410..23430  
/rpt\_family="AT-rich"  
23462..23519  
/rpt\_family="MALR"  
23600..23748  
/rpt\_family="MALR"  
23749..24334  
/rpt\_family="ERV1"  
24335..24548  
/rpt\_family="MALR"  
24600..24700  
/rpt\_family="MALR"  
24975..25812  
/rpt\_family="L2"  
BASE COUNT 7711 a 5089 c 5453 g 7616 t  
ORIGIN

Query Match 45.0%; Score 28.8; DB 9; Length 25869;  
 Best Local Similarity 65.6%; Pred. No. 18;  
 Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Db 19526 CAAAAATACCTGCGCTTGAAGCTTAATAATACATCCCGCCCGCATGTTCCATCA 19467

Db 19466 AGCC 19463

RESULT 14  
 AC129514/c 66873 bp DNA linear HTG 30-JUL-2002  
 LOCUS Homo sapiens chromosome 18 clone RP11-61K3 map 18, LOW-PASS  
 DEFINITION  
 AC129514  
 AC129514.1 GI:22004327  
 VERSION  
 KEYWORDS  
 HTG; HTGS\_PPHASE0.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

1 (bases 1 to 66873)  
 Bliren, B., Nussbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barre, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalil, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fair, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
 Gaidyna, S., Gird, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
 Horton, L., Huime, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
 Karatas, A., Kellis, C., Landers, T., Major, J., Matthews, C.,  
 Liu, G., Maclean, C., Macdonald, P., May, J., Mhova, T., Mlenga, V.,  
 McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Petersen, K.,  
 Punukhang, P., Piere, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Testaye, S., Theodore, J., Topman, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L27809  
 Center clone name: 61\_K\_3

NOTE: This record contains 81 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that

\* the record is updated, the accession number will  
 \* be preserved.

1 729: contig of 729 bp in length  
 730 829: gap of 100 bp  
 830 1552: contig of 723 bp in length  
 1553 1652: gap of 100 bp  
 1653 2377: contig of 725 bp in length  
 2378 2477: gap of 100 bp  
 2478 3209: contig of 732 bp in length  
 3210 3309: gap of 100 bp  
 3310 4045: contig of 736 bp in length  
 4046 4145: gap of 100 bp  
 4146 4868: contig of 723 bp in length  
 4869 4968: gap of 100 bp  
 4969 5703: contig of 735 bp in length  
 5704 5803: gap of 100 bp  
 5804 6549: contig of 746 bp in length  
 6550 6649: gap of 100 bp  
 6650 7387: contig of 738 bp in length  
 7388 7487: gap of 100 bp  
 7488 8212: contig of 725 bp in length  
 8213 8312: gap of 100 bp  
 8313 9033: contig of 721 bp in length  
 9034 9133: gap of 100 bp  
 9134 9942: gap of 100 bp  
 9943 10672: contig of 730 bp in length  
 10673 10772: gap of 100 bp  
 10773 11514: contig of 742 bp in length  
 11515 11614: gap of 100 bp  
 11615 12328: contig of 714 bp in length  
 12329 12428: gap of 100 bp  
 12429 13168: contig of 740 bp in length  
 13169 13288: gap of 100 bp  
 13289 13992: contig of 724 bp in length  
 13993 14092: gap of 100 bp  
 14093 14819: contig of 727 bp in length  
 14820 14919: gap of 100 bp  
 14920 15646: contig of 727 bp in length  
 15647 15746: gap of 100 bp  
 15747 16470: contig of 724 bp in length  
 16471 16570: gap of 100 bp  
 16571 17299: contig of 729 bp in length  
 17300 17399: gap of 100 bp  
 17400 18129: contig of 730 bp in length  
 18130 18229: gap of 100 bp  
 18230 18958: contig of 729 bp in length  
 18959 19058: gap of 100 bp  
 19060 19774: contig of 716 bp in length  
 19775 19874: gap of 100 bp  
 19875 20611: contig of 737 bp in length  
 20612 20711: gap of 100 bp  
 20712 21448: contig of 737 bp in length  
 21449 21548: gap of 100 bp  
 21549 22300: contig of 752 bp in length  
 22301 22400: gap of 100 bp  
 22401 23141: contig of 741 bp in length  
 23142 23241: gap of 100 bp  
 23242 23966: contig of 725 bp in length  
 23967 24066: gap of 100 bp  
 24067 24788: contig of 722 bp in length  
 24789 24888: gap of 100 bp  
 24889 25604: contig of 716 bp in length  
 25605 25704: gap of 100 bp  
 25705 26430: contig of 726 bp in length  
 26431 26530: gap of 100 bp  
 26531 27252: contig of 722 bp in length  
 27253 27352: gap of 100 bp  
 27353 28083: contig of 731 bp in length  
 28084 28183: gap of 100 bp  
 28184 28887: contig of 704 bp in length  
 28888 28987: gap of 100 bp  
 28988 29731: contig of 744 bp in length





Fri May 30 09:39:40 2003

Search completed: May 5, 2003, 19:24:44  
Job time : 557.297 secs

us-09-990-099-23.rge

---



PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 29473; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 815 BP; 210 A; 203 C; 187 G; 215 T; 0 other;

Query Match 100.0%; Score 64; DB 23; Length 815;  
Best Local Similarity 100.0%; Pred. No. 4.4e-13;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAAGAAAGTCCGCTGATAGGCTTGAAGTTCATTTCCAGACCATTTTACATCGT 60  
DB 100 CAAAGAAAGTCCGCTGATAGGCTTGAAGTTCATTTCCAGACCATTTTACATCGT 159  
61 AGCC 64  
1111  
DB 160 AGCC 163

RESULT 2  
AA194502/c  
ID AA194502 standard; cDNA; 758 BP.

XX AA194502;

DT 13-NOV-2001 (first entry)

DE Human neuroblastoma expressed polynucleotide SEQ ID NO 577.

KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

OS Homo sapiens.

PN WO200166719-A1.

PD 13-SEP-2001.

PF 02-MAR-2001; 2001WO-JP01629.

PR 07-MAR-2000; 2000JP-0159195.

PA (CHIB-) CHIBA PREFECTURE.

PA (HISM) HISAMITSU PHARM CO LTD.

PI Nakagawara A;

DR WPI; 2001-565584/63.

PT Nucleic acids originating in gene expressed in human neuroblastoma,

PT useful as probe or primer in diagnosing prognosis of human  
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker  
PT for anti-cancer agents

PS Claim 1; Page 464; 2979pp; Japanese.

XX The invention relates to novel genes (AA193926-AA197963) expressed in  
CC human neuroblastoma. The nucleic acids are applicable as a probe or  
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
CC susceptibility indicators or tumour markers for anti-cancer agents. The  
CC gene information for diagnosing prognosis is related to factors similar  
CC to that for N-myc and TrkA genes.

SO Sequence 758 BP; 193 A; 162 C; 180 G; 194 T; 29 other;

Query Match 44.4%; Score 28.4; DB 22; Length 758;  
Best Local Similarity 64.4%; Pred. No. 1.7;  
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 3 AAAAAAGTCCGCTGATAGGCTTGAAGTTCATTTCCAGACCATTTTACATCGTA 61  
DB 711 AAGGACGCTCCGNGTTANANTGAACAATTAATTTCCANNACATTTACGTTGTA 653

RESULT 3  
AAS71401  
ID AAS71401 standard; cDNA; 1419 BP.

XX AAS71401;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #7205.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABC07214.

PS New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 1; SEQ ID No 7205; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful for treating  
CC imaging of sites expressing (II). (I) and (II) are useful for treating





DT 09-OCT-2001 (first entry)

PR 23-MAR-2000; 2000US-191637P  
PR 11-JUL-2000; 2000US-0614150

RESULT 7	
AAH92117/c	
ID	AAH92117 standard; DNA; 700 BP
xx	
AC	AAH92117;
xx	
DDT	09-OCT-2001 (first entry)
xx	

XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, LI PWD, Myers EW;  
 XX WPI: 2001-656860/75.  
 DR New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1: SEQ ID NO 32893; 21pp + Sequence listing: English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 2363 BP; 729 A; 471 C; 490 G; 673 T; 0 other:  
 SQ  
 XX  
 XX Query Match 38.8%; Score 24.8; DB 23; Length 2363;  
 XX Best Local Similarity 72.7%; Pred. No. 40;  
 XX Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 OY 11 TCCGCTGATTAAGCTTGAAGTTTCATTTCCACACCACTTTTA 54  
 DB 591 TCCGATGATTAATTAATTCAGAGTTTCATTTCCACACCACTTTTA 548  
 ID ABL27142 standard; DNA: 4001 BP.  
 XX  
 XX ABL27142;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 32899.  
 DE  
 XX Drosophila; developmental biology; cell signalling; insecticides;  
 KW pharmaceutical; gene; ds.  
 KW  
 XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PE  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PR  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, LI PWD, Myers EW;  
 PI  
 XX WPI: 2001-656860/75.  
 DR  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1: SEQ ID NO 32899; 21pp + Sequence listing: English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 4001 BP; 1135 A; 799 C; 940 G; 1127 T; 0 other:  
 SQ  
 XX  
 XX Query Match 38.8%; Score 24.8; DB 23; Length 4001;  
 XX Best Local Similarity 72.7%; Pred. No. 45;  
 XX Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 OY 11 TCCGCTGATTAAGCTTGAAGTTTCATTTCCACACCACTTTTA 54  
 DB 3152 TCCGATGATTAATTAATTCAGAGTTTCATTTCCACACCACTTTTA 3109  
 ID AAS44601 standard; DNA: 1484 BP.  
 XX  
 XX AAS44601;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Human full-length polynucleotide sequence #26.  
 DE  
 XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;  
 KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
 KW nervous system disorder; inflammatory disorder; cell differentiation; ds;  
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
 KW cytoskeletal; antineoplastic; antiarthritic; vulnery; antiinflammatory;  
 KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
 KW neuroprotective; osteopathic; antidiabetic; antiallergic; antiallergic;  
 KW immunostimulant; analgesic; gene therapy.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200164834-A2.  
 PN  
 XX 07-SEP-2001.  
 PD  
 XX 26-FEB-2001; 2001WO-US04926.  
 PE  
 XX 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 PR 17-JUN-2000; 2000US-0597707.  
 PR 14-JUL-2000; 2000US-0616807.  
 PR 19-SEP-2000; 2000US-0664641.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
 PI Drmanac R;  
 PI  
 XX WPI: 2001-589862/66.  
 DR P-PSDB; AAU27701.  
 DR  
 XX Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis, treatment of  
 PT cancer, neurological, inflammatory disorders and for use in arrays for  
 PT detection -  
 XX  
 XX Claim 1: SEQ ID NO 26; 153pp; English.

CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and  
 CC contig polynucleotides encoding polypeptides of the invention. The DNA  
 CC and protein sequences are useful for the treatment, diagnosis and  
 CC prevention of various types of disorder in a mammalian subject such as a  
 CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers  
 CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such  
 CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's  
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory  
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
 CC cell proliferation, cell differentiation, stem cell growth factor,  
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
 CC in culture to give rise to neuroepithelial cells that can be used to  
 CC augment or replace cells damaged by illness, accidental damage or genetic  
 CC disorders. The sequences may also be used for regeneration of bone,  
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
 CC Note: Some sequences for this patent did not form part of the printed  
 CC specification, but were obtained in electronic format directly from WIPO  
 CC at ftp://wipo.int/pub/published\_pcr\_sequences.

SO Sequence 1484 BP: 448 A: 297 C: 293 G: 446 T: 0 other:

Query Match 38.4% Score 24.6; DB 22; Length 1484;  
 Best Local Similarity 61.9%; Pred. No. 43;  
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 AAAAAAAGTCGGTGAATAGCGCTTGAAGAAAGTTCATTCACAGCCATTTCATGATGTA 61  
 DB 120 AAAAAAAGTCGGTGAATAGCGCTTGAAGAAAGTTCATTCATTCATTCATTCATGTA 179

QY 62 GCC 64  
 DB 180 GGC 182

RESULT 11  
 AAS44601/c  
 ID AAS44601 standard; DNA: 1484 BP.

AC AAS44601;

DT 18-DEC-2001 (first entry)

DE Human full-length polynucleotide sequence #26.

XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;  
 KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
 KW nervous system disorder; inflammatory disorder; cell differentiation; ds;  
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
 KW cytoskeletal; antiinflammatory; antiarthritic; vulnery; antiinflammatory;  
 KW antibacterial; immunosuppressive; vasculoprotective; antiparkinsonian;  
 KW neuroprotective; osteoprotective; antidiabetic; antiallergic;  
 KW immunostimulant; analgesic; gene therapy.

OS Homo sapiens.

PN WO200164834-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04926.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PR 17-JUN-2000; 2000US-0597707.

PR 14-JUL-2000; 2000US-0616807.

PR 19-SEP-2000; 2000US-0664641.

PA (HYSE-) HYSEQ INC.

PI Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
 PI Dynamic R;

DR WPI: 2001-589862/66.  
 DR P-PSDB: AAU27701.

PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis, treatment of  
 PT cancer, neurological, inflammatory disorders and for use in arrays for  
 PT detection

XX Claim 1; SEQ ID No 26; 153bp; English.

CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and  
 CC contig polynucleotides encoding polypeptides of the invention. The DNA  
 CC and protein sequences are useful for the treatment, diagnosis and  
 CC prevention of various types of disorder in a mammalian subject such as a  
 CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers  
 CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such  
 CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's  
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory  
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
 CC cell proliferation, cell differentiation, stem cell growth factor,  
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
 CC in culture to give rise to neuroepithelial cells that can be used to  
 CC augment or replace cells damaged by illness, accidental damage or genetic  
 CC disorders. The sequences may also be used for regeneration of bone,  
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
 CC Note: Some sequences for this patent did not form part of the printed  
 CC specification, but were obtained in electronic format directly from WIPO  
 CC at ftp://wipo.int/pub/published\_pcr\_sequences.

SO Sequence 1484 BP: 448 A: 297 C: 293 G: 446 T: 0 other:

Query Match 38.4% Score 24.6; DB 22; Length 1484;  
 Best Local Similarity 61.9%; Pred. No. 43;  
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 AAAAAAAGTCGGTGAATAGCGCTTGAAGAAAGTTCATTCACAGCCATTTCATGATGTA 61  
 DB 1369 AAAAAAAGTCGGTGAATAGCGCTTGAAGAAAGTTCATTCATTCATTCATTCATGTA 1310

QY 62 GCC 64  
 DB 1309 GGC 1307

RESULT 12

AAV41728/c  
 ID AAV41728 standard; DNA: 1865 BP.

XX AAV41728;

DT 20-NOV-1998 (first entry)

DE DNA encoding the mature protein of human serum albumin (HSA).

XX Protein expression; monocotyledon plant cell;

KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;  
 KW AATII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;  
 KW antithrombotic; blood replacement; ss.

OS Homo sapiens.

PN WO9836085-A1.

PD 20-AUG-1998.  
 XX 13-FEB-1998; 98WO-US03068.  
 PF 13-FEB-1997; 97US-0038170.  
 PR 13-FEB-1997; 97US-0037991.  
 PR 13-FEB-1997; 97US-0038168.  
 PR 13-FEB-1997; 97US-0038169.  
 XX (PHT-) APPLIED PHYTOLOGICS INC.  
 PA Rodriguez RL, Sutliff TD;  
 XX WPI: 1998-467179/40.  
 DR P-PSDB: AAM59841.  
 XX  
 XX Expressing mature, glycosylated proteins in monocotyledonous plant  
 PT cells - from chimeric gene including signal peptide sequence,  
 PT specifically therapeutic agents and industrial enzymes  
 XX  
 PS Disclosure: Pages 32-33; 53pp; English.  
 XX  
 XX The present sequence encodes the mature protein of human serum albumin  
 CC (HSA). The protein is used to exemplify the invention. The  
 CC specification describes a method for producing mature heterologous  
 CC protein in monocotyledonous plant cells. The method comprises  
 CC transforming the cells with a chimeric gene comprising a monocotyledon  
 CC transcription regulator, inducible either during seed maturation or by  
 CC adding/removing a small molecule, DNA encoding the heterologous protein,  
 CC and DNA encoding a signal peptide, with the signal peptide causing  
 CC secretion of the protein from the cell. Proteins expressed in this  
 CC manner include mature glycosylated alpha 1-antitrypsin (AAT) with a  
 CC glycosylation pattern that significantly increases its serum half-life,  
 CC mature glycosylated antithrombin III (ATIII), mature human serum albumin  
 CC (HSA) having the native folding pattern as shown by bilirubin-binding  
 CC characteristics, or mature active subtilisin BPN'. These proteins are  
 CC useful therapeutically (e.g. AAT for treating emphysema, ATIII as  
 CC antithrombotic and HSA as blood replacement) or as industrial enzymes  
 CC (BPN' is used in detergents).  
 CC  
 XX Sequence 1865 BP; 587 A; 369 C; 422 G; 487 T; 0 other;  
 SQ  
 Query Match 38.4%; Score 24.6; DB 19; Length 1865;  
 Best Local Similarity 61.9%; Pred. No. 45;  
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 2 AAAAAAAAAAGTCGCTGATAGCTTGAAGTTCATTTCAGACCATTTTACATGTA 61  
 DB 1845 AAAAAAGAACAGATAGATAGCTTTCATCTTCCTTCATTTCATGTA 1786  
 OY 62 GCC 64  
 DB 1785 GGC 1783  
 XX  
 XX RESULT 13  
 XX AAV41732/C  
 XX ID AAV41732 standard; DNA: 1940 BP.  
 XX  
 XX AAV41732;  
 XX  
 XX 20-NOV-1998 (first entry)  
 XX  
 XX Codon-optimised Ramy3D signal fused to DNA encoding mature HSA.  
 XX  
 XX Protein expression; monocotyledon plant cell;  
 XX glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;  
 XX ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;  
 XX antithrombotic; blood replacement; ss.  
 XX  
 XX Synthetic.  
 XX OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT misc\_feature 1..75  
 FT misc\_feature /\*tag= a  
 FT misc\_feature /note= "codon-optimised Ramy3D signal sequence"  
 FT misc\_feature 76..1940  
 FT misc\_feature /\*tag= b  
 FT /note= "encodes mature HSA"  
 XX  
 XX W09836085-A1.  
 XX  
 XX 20-AUG-1998.  
 XX  
 XX 13-FEB-1998; 98WO-US03068.  
 XX  
 XX 13-FEB-1997; 97US-0038170.  
 XX 13-FEB-1997; 97US-0037991.  
 XX 13-FEB-1997; 97US-0038168.  
 XX 13-FEB-1997; 97US-0038169.  
 XX (PHT-) APPLIED PHYTOLOGICS INC.  
 PA Rodriguez RL, Sutliff TD;  
 XX WPI: 1998-467179/40.  
 DR  
 XX  
 XX Expressing mature, glycosylated proteins in monocotyledonous plant  
 PT cells - from chimeric gene including signal peptide sequence,  
 PT specifically therapeutic agents and industrial enzymes  
 XX  
 PS Disclosure: Pages 34 iv-v; 53pp; English.  
 XX  
 XX The present sequence encodes a fusion protein of codon-optimised Ramy3D  
 CC signal sequence/mature human serum albumin (HSA). The protein is used  
 CC to exemplify the invention. The specification describes a method for  
 CC producing mature heterologous protein in monocotyledonous plant cells.  
 CC The method comprises transforming the cells with a chimeric gene  
 CC comprising a monocotyledon transcription regulator, inducible either  
 CC during seed maturation or by adding/removing a small molecule, DNA  
 CC encoding the heterologous protein, and DNA encoding a signal peptide,  
 CC with the signal peptide causing secretion of the protein from the cell.  
 CC Proteins expressed in this manner include mature glycosylated alpha  
 CC 1-antitrypsin (AAT) with a glycosylation pattern that significantly  
 CC increases its serum half-life, mature glycosylated antithrombin III  
 CC (ATIII), mature human serum albumin (HSA) having the native folding  
 CC pattern as shown by bilirubin-binding characteristics, or mature active  
 CC subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for  
 CC treating emphysema, ATIII as antithrombotic and HSA as blood replacement)  
 CC or as industrial enzymes (BPN' is used in detergents).  
 CC  
 XX Sequence 1940 BP; 598 A; 401 C; 438 G; 503 T; 0 other;  
 SQ  
 Query Match 38.4%; Score 24.6; DB 19; Length 1940;  
 Best Local Similarity 61.9%; Pred. No. 45;  
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 2 AAAAAAAAAAGTCGCTGATAGCTTGAAGTTCATTTCAGACCATTTTACATGTA 61  
 DB 1920 AAAAAAGAACAGATAGATAGCTTTCATCTTCCTTCATTTCATGTA 1861  
 OY 62 GCC 64  
 DB 1860 GGC 1858  
 XX  
 XX RESULT 14  
 XX AAN60107/C  
 XX ID AAN60107 standard; cDNA: 2015 BP.  
 XX  
 XX AAN60107;  
 XX  
 XX 29-JUN-1991 (first entry)  
 XX  
 XX Sequence encoding prepro human serum albumin (HSA).  
 XX

```

XX KW Shock therapy: hypoproteinaemia; erythroblastosis foetalis;
XX KW burn therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers.
XX FT CDS 32..85 /*tag= a
XX FT CDS /*product= pre sequence
XX FT CDS 86..103 /*tag= b
XX FT CDS /*product= pro sequence
XX FT mat_peptide 104..1861 /*tag= c
XX PN EP206733-A.
XX PD 30-DEC-1986.
XX PF 17-JUN-1986; 86EP-0304656.
XX PR 17-JUN-1985; 85US-0745524.
XX PR 17-JUN-1986; 86JP-0141206.
XX PR 19-DEC-1986; 86JP-0303613.
XX PA (GENE-) GENEX CORP.
XX PA (MITO ) MITSUBISHI DENKI KK.
XX PI Burns AL.
XX DR WPI: 1986-341032/52.
XX DR P-PSDB; AAP60092.
XX PT New synthetic human serum albumin gene - useful in plasmids for
XX PT transforming microorganisms
XX PS Disclosure; Fig 2; 44pp; English.
XX CC An E.coli HB101 culture transformed with the plasmid (PGX401)
XX CC containing the prepro-HSA gene (AAN60107) has been deposited with the
XX CC U.S. Department of Agriculture Northern Regional Research Laboratory
XX CC in Peoria, Illinois, as NRRL No. B-15784. PGX401 is specifically
XX CC claimed. It is believed that the AA sequence set forth in AAP60092
XX CC represents a genomic HSA allele that is widespread in the human
XX CC population. AAN60106 includes all variations of the HSA gene permitted
XX CC by the genetic code.
SQ Sequence 2015 BP; 612 A; 406 C; 447 G; 550 T; 0 other;

Query Match 38.4%; Score 24.6; DB 7; Length 2015;
Best Local Similarity 61.9%; Pred. NO. 46;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGTCGCTGATAGGCTTGAAGTTCATTTCCAGACCAATTTTTCATGTA 61
DB 1947 AAAAAAGAAACAGATGAATAAGCTTTGATGCTTCATTTCTTCTTATTCATGTA 1888
OY 62 GCC 64
DB 1887 GGC 1885

RESULT 15
AAN92583/C
ID AAN92583 standard; DNA; 2064 BP.
XX AC AAN92583;
XX DT 13-MAR-1992 (first entry)
XX DE Sequence of human serum albumin 13 (HSA 13).
XX

```

```

XX KW Albumin; yeast expression vector; methylotrophic yeast; ss.
XX OS Homo sapiens.
XX PN EP344459-A.
XX PD 06-DEC-1989.
XX PF 25-APR-1989; 89EP-0107459.
XX PR 25-APR-1988; 88US-0186420.
XX PA (PHIP ) PHILLIPS PETROLEUM CO.
XX PI Marashi F, Fuke M, Mcombie WR, Barr RD, Sreekishna K;
XX PI Wong R;
XX DR WPI: 1989-357919/49.
XX PT HSA prodn. in yeast - comprises methylotrophic yeast transformed
XX PT with vectors encoding HSA
XX PS Claim 24; Table 1, pages 4-5; 28pp; English.
XX CC The inventors claim the sequence in AAN92583, and vectors contg. a
XX CC structural HSA gene, a plasmid for expression in yeast and
XX CC methylotrophic yeast transformed by the vector. The vector
XX CC pHSAl13 and Pichia pastoris 95115/pHSAl13 are specifically claimed.
SQ Sequence 2064 BP; 649 A; 406 C; 448 G; 561 T; 0 other;

Query Match 38.4%; Score 24.6; DB 10; Length 2064;
Best Local Similarity 61.9%; Pred. NO. 46;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGTCGCTGATAGGCTTGAAGTTCATTTCCAGACCAATTTTTCATGTA 61
DB 1940 AAAAAAGAAACAGATGAATAAGCTTTGATGCTTCATTTCTTCTTATTCATGTA 1881
OY 62 GCC 64
DB 1880 GGC 1878

Search completed: May 5, 2003, 18:05:50
Job time : 440.637 secs

```

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:51:17 ; Search time 11.6044 Seconds

(Without alignments)  
1691.369 Million cell updates/sec

Title: US-09-990-099-23

Perfect score: 64  
Sequence: 1 caaaaaaaaaagtcgctgata.....cccatcttaccatcgtagcc 64

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued\_Patents.NA.\*

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCRTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Dackfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24.6	38.4	231	1	Sequence 8, Appl
2	24.6	38.4	231	1	Sequence 18, Appl
3	24.6	38.4	19011	1	Sequence 36, Appl
4	24.6	38.4	19557	5	Sequence 1, Appl
5	24.6	38.1	443	4	Sequence 180, App
6	23.4	36.6	4451	4	Sequence 45, Appl
7	23.4	36.6	4451	4	Sequence 45, Appl
8	23.4	36.6	4481	4	Sequence 1, Appl
9	23.4	36.6	4481	5	Sequence 6, Appl
10	23.4	36.6	4775	4	Sequence 37, Appl
11	23.4	36.6	4775	4	Sequence 37, Appl
12	23.4	36.6	4910	4	Sequence 40, Appl
13	23.4	36.6	4910	4	Sequence 48, Appl
14	23.4	36.6	5258	4	Sequence 48, Appl
15	23.4	36.6	5258	4	Sequence 48, Appl
16	23.4	36.6	5600	4	Sequence 3, Appl
17	23.4	36.6	5600	5	Sequence 8, Appl
18	23.2	36.2	3212	1	Sequence 1, Appl
19	23.2	36.2	10614	1	Sequence 35, Appl
20	23.2	36.2	10614	1	Sequence 35, Appl
21	23.2	35.9	2504	2	Sequence 4, Appl
22	23.2	35.9	2504	2	Sequence 4, Appl
23	22.6	35.3	2321	4	Sequence 2, Appl
24	22.6	35.3	2321	4	Sequence 2, Appl
25	22.6	35.3	2337	4	Sequence 3, Appl
26	22.6	35.3	2337	4	Sequence 3, Appl
27	22.4	35.0	8651	4	Sequence 181, App

28	22.2	34.7	75	1	US-07-770-100C-2	Sequence 2, Appl
29	22.2	34.7	1929	4	US-08-750-180-1	Sequence 1, Appl
30	22.2	34.7	43804	4	US-09-171-661-1	Sequence 7, Appl
31	22	34.4	3155	2	US-08-591-629-7	Sequence 209, App
32	21.8	34.1	1160	4	US-08-858-207A-209	Sequence 6, Appl
33	21.8	34.1	2004	1	US-08-471-033-6	Sequence 6, Appl
34	21.8	34.1	2004	2	US-08-471-044-6	Sequence 6, Appl
35	21.8	34.1	2004	2	US-08-463-483A-6	Sequence 6, Appl
36	21.8	34.1	2004	2	US-08-471-046A-6	Sequence 6, Appl
37	21.8	34.1	2004	2	US-08-470-566B-6	Sequence 6, Appl
38	21.8	34.1	2004	2	US-08-469-334-6	Sequence 6, Appl
39	21.8	34.1	2004	3	US-09-300-529-6	Sequence 6, Appl
40	21.8	34.1	2408	2	US-08-870-827-5	Sequence 5, Appl
41	21.8	34.1	2408	4	US-09-317-179-5	Sequence 5, Appl
42	21.8	34.1	2645	4	US-08-960-780-31	Sequence 31, Appl
43	21.8	34.1	2645	4	US-08-073-898-31	Sequence 1, Appl
44	21.8	34.1	2645	4	US-09-371-913A-1	Sequence 1, Appl
45	21.8	34.1	2655	1	US-08-471-033-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-07-691-079C-8/c  
Sequence 8, Application US/07691079C

Patent No. 5330301

GENERAL INFORMATION:

APPLICANT: William D. Prevatt et al.

TITLE OF INVENTION: Expression of Human Serum Albumin in

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: RICHMOND, PHILLIPS, HITCHCOCK & UMPHRETT

STREET: P. O. Box 2443

CITY: Bartlesville

STATE: OK.

COUNTRY: USA

ZIP: 74005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM PC

SOFTWARE: Display Write 4

OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICATION NUMBER: US/07/691,079C

FILING DATE: 19910426

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hal Brent Woodrow

REGISTRATION NUMBER: 32,501

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-918-661-0624

INFORMATION FOR SEQ. ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 231bp

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Linker Oligonucleotide

US-07-691-079C-8

Query Match

Best local Similarity 61.9%;

Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Score 24.6; DB 1; Length 231;

Pred. No. 3.4;

2 AAAAAAGTCGCTGTAAGCTTGAAGCTTCCAGACCATTTTACATCGTA 61

102 AAAAAACACAGATGATTAAGCTTTCATTTCTTCTATTCATCGATGA 43

62 GCC 64

CORRESPONDENCE

LOCATION: 6802..6934  
OTHER INFORMATION: /number= 5



FEATURE:  
NAME/KEY: intron  
LOCATION: 6935..7758  
OTHER INFORMATION: /number= 5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 7759..7856  
OTHER INFORMATION: /number= 6  
FEATURE:  
NAME/KEY: intron  
LOCATION: 7857..9443  
OTHER INFORMATION: /number= 6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 9444..9573  
OTHER INFORMATION: /number= 7  
FEATURE:  
NAME/KEY: intron  
LOCATION: 9574..10866  
OTHER INFORMATION: /number= 7  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10867..11081  
OTHER INFORMATION: /number= 8  
FEATURE:  
NAME/KEY: intron  
LOCATION: 11082..12480  
OTHER INFORMATION: /number= 8  
FEATURE:  
NAME/KEY: exon  
LOCATION: 12481..12613  
OTHER INFORMATION: /number= 9  
FEATURE:  
NAME/KEY: intron  
LOCATION: 12614..13701  
OTHER INFORMATION: /number= 9  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13702..13799  
OTHER INFORMATION: /number= 10  
FEATURE:  
NAME/KEY: intron  
LOCATION: 13800..14976  
OTHER INFORMATION: /number= 10  
FEATURE:  
NAME/KEY: exon  
LOCATION: 14977..15115  
OTHER INFORMATION: /number= 11  
FEATURE:  
NAME/KEY: intron  
LOCATION: 15116..15533  
OTHER INFORMATION: /number= 11  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15534..15757  
OTHER INFORMATION: /number= 12  
FEATURE:  
NAME/KEY: intron  
LOCATION: 15758..16949  
OTHER INFORMATION: /number= 12  
FEATURE:  
NAME/KEY: exon  
LOCATION: 16950..17082  
OTHER INFORMATION: /number= 13  
FEATURE:  
NAME/KEY: intron  
LOCATION: 17083..17696  
OTHER INFORMATION: /number= 13  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17697..17764  
OTHER INFORMATION: /number= 14  
FEATURE:

NAME/KEY: intron  
LOCATION: 17765..18534  
OTHER INFORMATION: /number= 14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1776..1854, 2564..2621, 4076..4208,  
6041..6252, 6802..6934, 7759..7856, 9444..9573,  
10867..11081, 12481..12613, 13702..13799,  
14977..15115, 15534..15757, 16950..17082,  
17697..17741)  
OTHER INFORMATION: /product= "human serum albumin"  
OTHER INFORMATION: /citation= ([1])  
FEATURE:  
NAME/KEY: exon  
LOCATION: 18535..18697  
OTHER INFORMATION: /number= 15  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 17742..18697  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1737..1775  
PUBLICATION INFORMATION:  
AUTHORS: Mungrettl, P P  
AUTHORS: Ruffner, D E  
AUTHORS: Kuang, W-J  
AUTHORS: Dennison, O E  
AUTHORS: Hawkins, J W  
AUTHORS: Beattie, W G  
AUTHORS: Dugaiczky, A  
TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN  
TITLE: GENE IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN  
JOURNAL: J. Biol. Chem.  
VOLUME: 261  
PAGES: 6747-6757  
DATE: 1986  
RELEVANT RESIDUES IN SEQ ID NO: 36: FROM 1 TO 19011  
US-08-310-356-36  
Query Match 38.4%; Score 24.6; DB 1; Length 19011;  
Best Local Similarity 61.9%; Pred. No. 9.5;  
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 2 AAAAAAGTCCGCTGTAAGCTTGAAGTTCATTCAGCCATTTTACATCGTA 61  
Db 18596 AAAAAAGTCCGCTGTAAGCTTGAAGTTCATTCATTCATTCATTCATGTA 18537  
QY 62 GCC 64  
Db 18536 GCC 18534  
RESULT 4  
PCT-US92-06300-1/c  
Sequence 1, Application PC/TUS9206300  
GENERAL INFORMATION:  
APPLICANT: Hurwitz, David R  
APPLICANT: Nathan, Margaret  
TITLE OF INVENTION: Transgenic Protein Production  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer, Inc.  
STREET: 500 Virginia Ave., Bldg. 3A  
CITY: Ft. Washington  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19034  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06300  
FILING DATE: 19920730  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Rosanne  
REGISTRATION NUMBER: 52,534  
REFERENCE/DOCKET NUMBER: A0856-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 962-4130  
TELEFAX: (215) 962-4107  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1957 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Minghelli, P P  
AUTHORS: Ruffner, D E  
AUTHORS: Kuang, W.-J.  
AUTHORS: Dennison, O E  
AUTHORS: Hawkins, J W  
AUTHORS: Beattie, W G  
AUTHORS: Dugalczyk, A  
TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE  
TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22  
JOURNAL: J. Biol. Chem.  
VOLUME: 261  
PAGES: 6747-6757  
DATE: 1986  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002  
PCT-US92-06300-1

Query Match  
Best Local Similarity 61.9%; Pred. No. 9.6;  
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 AAAAAAGCTCCGCTGTAAGCTTGAAAAGTTGATTCAGACCCATTTTACATGTA 61  
DB 19142 AAAAAAGAAAGAGATGATTAAGCTTTGATCTTCATTTCTTCTCTATTCATGTA 19083

QY 62 GCC 64  
DB 19082 GCC 19080

RESULT 5  
US-08-936-165A-180  
Sequence 180, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nichols, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Rosenber, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic polynucleotides,  
TITLE OF INVENTION: Polypeptides and their uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road

CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 180:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 443 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-180

Query Match  
Best Local Similarity 73.8%; Pred. No. 4.7;  
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 16 TGATAGGCTTGAAAGTTGATTCAGACCCATTTTACAT 57  
DB 200 TGATTAGGTTCAAGTTGATTTTCACACATCTAATAT 241

RESULT 6  
US-09-303-064-45/c  
Sequence 45, Application US/09303064  
Patent No. 6221619  
GENERAL INFORMATION:  
APPLICANT: MAINE, Gregory T.  
APPLICANT: HUNT, Jeffery C.  
APPLICANT: BROJANAC, Susan  
APPLICANT: JYH-TSING SHEU, Michael  
APPLICANT: CHOVAN, Linda E.  
APPLICANT: TYNER, Joan D.  
APPLICANT: HOWARD, Lawrence V.  
APPLICANT: PARMELEY, Stephen F.  
APPLICANT: REMINGTON, Jack S.  
APPLICANT: ARAUJO, Fausto  
APPLICANT: SUZUKI, Yasuhiko  
APPLICANT: LI, Shuli  
TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF  
FILE REFERENCE: 6361 US, P1  
CURRENT APPLICATION NUMBER: US/09/303,064  
CURRENT FILING DATE: 1999-04-30  
EARLIER APPLICATION NUMBER: 09/086,503  
EARLIER FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 45  
LENGTH: 4451  
TYPE: DNA  
ORGANISM: Toxoplasma gondii  
US-09-303-064-45

Query Match 36.6%; Score 23.4; DB 4; Length 4451;  
Best Local Similarity 73.2%; Pred. No. 18;  
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAAAGTCGCTGATAGCTTGAAAGTTCAATTCGA 42  
DB 1686 AAAAAAAGCCGCTCATATTAGCGGCTAAGCTTCTTGCCA 1646

RESULT 7  
US-09-086-503-45/C

; Sequence 45, Application US/09086503A  
; Patent No. 6329157  
; GENERAL INFORMATION:  
; APPLICANT: MAINE, Gregory T.  
; APPLICANT: HUNT, Jeffery C.  
; APPLICANT: BROJANAC, Susan  
; APPLICANT: JYH-TSING SHEU, Michael  
; APPLICANT: CHOYAN, Linda E.  
; APPLICANT: TYNER, Joan D.  
; APPLICANT: HOWARD, Lawrence V.  
; TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF  
; FILE REFERENCE: 6361.US.01  
; CURRENT APPLICATION NUMBER: US/09/086,503A  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 45  
; LENGTH: 4451  
; TYPE: DNA  
; ORGANISM: Toxoplasma gondii  
US-09-086-503-45

Query Match 36.6%; Score 23.4; DB 4; Length 4451;  
Best Local Similarity 73.2%; Pred. No. 18;  
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAAAGTCGCTGATAGCTTGAAAGTTCAATTCGA 42  
DB 1686 AAAAAAAGCCGCTCATATTAGCGGCTAAGCTTCTTGCCA 1646

RESULT 8  
US-08-867-611-1/C

; Sequence 1, Application US/08867611  
; Patent No. 6172189  
; GENERAL INFORMATION:  
; APPLICANT: DEVARE, SUSHIL G  
; APPLICANT: DESAI, SURESH M  
; APPLICANT: CASEY, JAMES M  
; APPLICANT: DAILEY, STEPHEN H  
; APPLICANT: DAMSON, GEORGE J  
; APPLICANT: GUTIERREZ, ROBIN A  
; APPLICANT: LESNIEWSKI, RICHARD R  
; APPLICANT: STEWART, JAMES L  
; APPLICANT: RUPPRECHT, KEVIN R  
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT  
; TITLE OF INVENTION: ANTIGENS  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES  
; STREET: ONE ABBOTT PARK ROAD, CHAND377/AP602  
; CITY: ABBOTT PARK  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/867,611

FILING DATE: 02-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/646,757

FILING DATE: US/08/179,896

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/572,822

FILING DATE: 24-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/614,069

FILING DATE: 07-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,561

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,565

FILING DATE: 21-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/748,566

FILING DATE: 21-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: FOREMBSKI, PRISCILLA E

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 4834.US.P6

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365

TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4481 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 130..1317

US-08-867-611-1

Query Match 36.6%; Score 23.4; DB 4; Length 4481;  
Best Local Similarity 73.2%; Pred. No. 18;  
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAAAGTCGCTGATAGCTTGAAAGTTCAATTCGA 42  
DB 1716 AAAAAAAGCCGCTCATATTAGCGGCTAAGCTTCTTGCCA 1676

RESULT 9

PCT-US92-06965A-6/C

; Sequence 6, Application PC/TUS9206965A  
; GENERAL INFORMATION:  
; APPLICANT: DEVARE, S.  
; APPLICANT: DESAI, S.  
; APPLICANT: DAILEY, S.  
; TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES  
; STREET: ONE ABBOTT PARK ROAD  
; CITY: ABBOTT PARK  
; STATE: ILLINOIS  
; COUNTRY: U.S.  
; ZIP: 60065-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06965A

```

; FILING DATE: 19920821
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834PC.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: circular
; FEATURE:
; MOLECULE TYPE: DNA (genomic)
; NAME/KEY: CDS
; LOCATION: 130..1317
; PCT-US92-06965A-6

```

Query Match 36.6%; Score 23.4; DB 5; Length 4481;

Best Local Similarity 73.2%; Pred. No. 18;

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

QY 2 AAAAAAAGTCGCTGATTAAGCTTGAAGTTCATTGCCA 42
   ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 AAAAAAAGCCGCTCATTAAGCGGCGCTAAGCTTCTGCCA 1676

```

RESULT 10

US-09-303-064-37/c

; Sequence 37, Application US/09303064

; Patent No. 6221619

; GENERAL INFORMATION:

; APPLICANT: MAINE, Gregory T.

; APPLICANT: HUNT, Jeffrey C.

; APPLICANT: BROJANAC, Susan

; APPLICANT: JYH-TSING SHEU, Michael

; APPLICANT: CHOYAN, Linda E.

; APPLICANT: TYNER, Joan D.

; APPLICANT: HOWARD, Lawrence V.

; APPLICANT: PARMELEY, Stephen F.

; APPLICANT: REMINGTON, Jack S.

; APPLICANT: ARAUJO, Fausto

; APPLICANT: SUZUKI, Yashuhiro

; APPLICANT: LI, Shuli

; TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF

; FILE REFERENCE: 6361.US.P1

; CURRENT APPLICATION NUMBER: US/09/303,064

; EARLIER FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 09/086,503

; NUMBER OF SEQ ID NOS: 55

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 37

; LENGTH: 4775

; TYPE: DNA

; ORGANISM: Toxoplasma gondii

US-09-303-064-37

Query Match 36.6%; Score 23.4; DB 4; Length 4775;

Best Local Similarity 73.2%; Pred. No. 18;

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

QY 2 AAAAAAAGTCGCTGATTAAGCTTGAAGTTCATTGCCA 42
   ||||| ||||| ||||| ||||| ||||| |||||
Db 2010 AAAAAAAGCCGCTCATTAAGCGGCGCTAAGCTTCTGCCA 1970

```

RESULT 11

US-09-086-503-37/c

; Sequence 37, Application US/09086503A

```

; Patent No. 6329157
; GENERAL INFORMATION:
; APPLICANT: MAINE, Gregory T.
; APPLICANT: HUNT, Jeffrey C.
; APPLICANT: BROJANAC, Susan
; APPLICANT: JYH-TSING SHEU, Michael
; APPLICANT: CHOYAN, Linda E.
; APPLICANT: TYNER, Joan D.
; APPLICANT: HOWARD, Lawrence V.
; TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
; FILE REFERENCE: 6361.US.01
; CURRENT APPLICATION NUMBER: US/09/086,503A
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; US-09-086-503-37

```

Query Match 36.6%; Score 23.4; DB 4; Length 4775;

Best Local Similarity 73.2%; Pred. No. 18;

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

QY 2 AAAAAAAGTCGCTGATTAAGCTTGAAGTTCATTGCCA 42
   ||||| ||||| ||||| ||||| ||||| |||||
Db 2010 AAAAAAAGCCGCTCATTAAGCGGCGCTAAGCTTCTGCCA 1970

```

RESULT 12

US-09-303-064-40/c

; Sequence 40, Application US/09303064

; Patent No. 6221619

; GENERAL INFORMATION:

; APPLICANT: MAINE, Gregory T.

; APPLICANT: HUNT, Jeffrey C.

; APPLICANT: BROJANAC, Susan

; APPLICANT: JYH-TSING SHEU, Michael

; APPLICANT: CHOYAN, Linda E.

; APPLICANT: TYNER, Joan D.

; APPLICANT: HOWARD, Lawrence V.

; APPLICANT: PARMELEY, Stephen F.

; APPLICANT: REMINGTON, Jack S.

; APPLICANT: ARAUJO, Fausto

; APPLICANT: SUZUKI, Yashuhiro

; APPLICANT: LI, Shuli

; TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF

; FILE REFERENCE: 6361.US.P1

; CURRENT APPLICATION NUMBER: US/09/303,064

; EARLIER FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 09/086,503

; NUMBER OF SEQ ID NOS: 55

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 40

; LENGTH: 4910

; TYPE: DNA

; ORGANISM: Toxoplasma gondii

US-09-303-064-40

Query Match 36.6%; Score 23.4; DB 4; Length 4910;

Best Local Similarity 73.2%; Pred. No. 19;

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

QY 2 AAAAAAAGTCGCTGATTAAGCTTGAAGTTCATTGCCA 42
   ||||| ||||| ||||| ||||| ||||| |||||
Db 2145 AAAAAAAGCCGCTCATTAAGCGGCGCTAAGCTTCTGCCA 2105

```

RESULT 13

US-09-086-503-40/c

; Sequence 40, Application US/09086503A

```
; Patent No. 6329157
; GENERAL INFORMATION:
; APPLICANT: MAINE, Gregory T.
; APPLICANT: HUNT, Jeffery C.
; APPLICANT: BROJANAC, Susan
; APPLICANT: JYH-TSING SHEU, Michael
; APPLICANT: CHOVAN, Linda E.
; APPLICANT: TYNER, Joan D.
; APPLICANT: HOWARD, Lawrence V.
; TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
; FILE REFERENCE: 6361.US.01
; CURRENT APPLICATION NUMBER: US/09/086,503A
; CURRENT FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 4910
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; US-09-086-503-40
```

```
Query Match          36.6%; Score 23.4; DB 4; Length 4910;
Best Local Similarity 73.2%; Pred. No. 19;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 2 AAAAAAGTCGCGTATAGGCTTGAAGTTCAATTCCCA 42
      ||||| ||||| ||||| ||||| ||||| |||||
DB 2145 AAAAAAGCCGCTCATTAAGCGGCGTAAAGCTTCTTGCCA 2105
```

```
RESULT 14
US-09-303-064-48/C
; Sequence 48, Application US/09303064
; Patent No. 6221619
; GENERAL INFORMATION:
; APPLICANT: MAINE, Gregory T.
; APPLICANT: HUNT, Jeffery C.
; APPLICANT: BROJANAC, Susan
; APPLICANT: JYH-TSING SHEU, Michael
; APPLICANT: CHOVAN, Linda E.
; APPLICANT: TYNER, Joan D.
; APPLICANT: HOWARD, Lawrence V.
; APPLICANT: PARMEY, Stephen F.
; APPLICANT: REMINGTON, Jack S.
; APPLICANT: ARAUJO, Fausto
; APPLICANT: SUZUKI, Yashuhiro
; APPLICANT: LI, Shuli
; TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
; FILE REFERENCE: 6361.US.P1
; CURRENT APPLICATION NUMBER: US/09/303,064
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 09/086,503
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 5258
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; US-09-303-064-48
```

```
Query Match          36.6%; Score 23.4; DB 4; Length 5258;
Best Local Similarity 73.2%; Pred. No. 19;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 2 AAAAAAGTCGCGTATAGGCTTGAAGTTCAATTCCCA 42
      ||||| ||||| ||||| ||||| ||||| |||||
DB 2493 AAAAAAGCCGCTCATTAAGCGGCGTAAAGCTTCTTGCCA 2453
```

```
RESULT 15
US-09-086-503-48/C
; Sequence 48, Application US/09086503A
```

```
; Patent No. 6329157
; GENERAL INFORMATION:
; APPLICANT: MAINE, Gregory T.
; APPLICANT: HUNT, Jeffery C.
; APPLICANT: BROJANAC, Susan
; APPLICANT: JYH-TSING SHEU, Michael
; APPLICANT: CHOVAN, Linda E.
; APPLICANT: TYNER, Joan D.
; APPLICANT: HOWARD, Lawrence V.
; TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
; FILE REFERENCE: 6361.US.01
; CURRENT APPLICATION NUMBER: US/09/086,503A
; CURRENT FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 5258
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; US-09-086-503-48
```

```
Query Match          36.6%; Score 23.4; DB 4; Length 5258;
Best Local Similarity 73.2%; Pred. No. 19;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 2 AAAAAAGTCGCGTATAGGCTTGAAGTTCAATTCCCA 42
      ||||| ||||| ||||| ||||| ||||| |||||
DB 2493 AAAAAAGCCGCTCATTAAGCGGCGTAAAGCTTCTTGCCA 2453
```

```
Search completed: May 5, 2003, 18:00:52
Job time : 15.6044 secs
```



Fri May 30 09:39:40 2003

us-09-990-099-23.rmpb

Page 1

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:52:42 ; Search time 29.3626 Seconds  
(without alignments)  
2575.509 Million cell updates/sec

Title: US-09-990-099-23  
Perfect score: 64  
Sequence: 1 caaataaagtcgctgata.....cccatcttcaatcgtacgc 64

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: 1: Published Applications, NA: \*  
2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq: \*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.8	41.9	445	10	US-09-783-590-762 Sequence 762, App
2	24.6	38.4	362	9	US-09-918-995-30115 Sequence 30115, A
3	24.6	38.4	394	9	US-09-918-995-7625 Sequence 7625, Ap
4	24.6	38.4	425	9	US-09-918-995-8119 Sequence 8119, Ap
5	24.6	38.4	468	9	US-09-918-995-31162 Sequence 31162, A
6	24.4	38.1	443	10	US-09-939-980-180 Sequence 180, App
7	24.4	38.1	826	10	US-09-822-830A-561 Sequence 561, App
8	24.4	38.1	1245	9	US-09-938-842A-5526 Sequence 5526, Ap
9	24.4	38.1	4923	7	US-08-781-986A-551 Sequence 752, App
10	24.2	37.8	1043	10	US-09-974-300-7252 Sequence 4354, Ap
11	24.2	37.8	1228	10	US-09-878-574-796 Sequence 796, App
12	23.8	37.2	380	10	US-09-815-242-6976 Sequence 10, Appl
13	23.8	37.2	1371	10	US-10-003-806-10 Sequence 45, Appl
14	23.6	36.9	173808	12	US-09-896-852-45 Sequence 37, Appl
15	23.4	36.6	4451	10	US-09-896-852-37 Sequence 40, Appl
16	23.4	36.6	4775	10	US-09-896-852-40 Sequence 48, Appl
17	23.4	36.6	4910	10	US-09-896-852-48 Sequence 8555, Ap
18	23.4	36.6	5258	10	US-09-896-852-48 Sequence 8555, Ap
19	23.2	36.2	337	9	US-09-796-692-8555 Sequence 8555, Ap

20	23.2	36.2	337	9	US-10-040-862-8555 Sequence 8555, Ap
21	23.2	36.2	436	9	US-09-918-995-9303 Sequence 9303, Ap
22	23.2	36.2	1161	7	US-08-781-986A-497 Sequence 497, App
23	23.2	36.2	2000	9	US-09-938-842A-4923 Sequence 4923, Ap
24	23.2	36.2	2742	10	US-09-815-242-3941 Sequence 3941, Ap
25	23.2	36.2	2802	10	US-09-815-242-6546 Sequence 6546, Ap
26	23.2	36.2	3779	10	US-09-925-300-712 Sequence 712, App
27	23.2	36.2	32768	9	US-09-070-927A-128 Sequence 128, App
28	23.2	36.2	1503841	10	US-09-946-807-1 Sequence 1, Appl1
29	23.2	36.2	1503841	10	US-09-795-668-1 Sequence 1, Appl1
30	23.2	36.2	1503841	10	US-09-795-668-1 Sequence 10655, A
31	23	35.9	343	10	US-09-867-701-10855 Sequence 710, App
32	23	35.9	449	10	US-09-770-444-710 Sequence 312, App
33	23	35.9	498	9	US-09-736-457-112 Sequence 312, App
34	23	35.9	498	9	US-09-902-941-312 Sequence 312, App
35	23	35.9	498	9	US-09-849-626-312 Sequence 1993, Ap
36	23	35.9	498	9	US-10-017-754-312 Sequence 4956, Ap
37	23	35.9	571	9	US-10-060-036-1993 Sequence 3893, Ap
38	23	35.9	711	10	US-09-974-300-4956 Sequence 4955, Ap
39	23	35.9	780	9	US-09-938-842A-3893 Sequence 397, App
40	23	35.9	789	10	US-09-974-300-4955 Sequence 397, App
41	23	35.9	2058	10	US-09-899-980A-4 Sequence 397, App
42	23	35.9	4044	12	US-10-014-927-18 Sequence 397, App
43	23	35.9	7470	9	US-10-073-961-397 Sequence 397, App
44	23	35.9	7470	9	US-10-073-961-398 Sequence 397, App
45	23	35.9	7470	10	US-09-764-887-397 Sequence 397, App

ALIGNMENTS

RESULT 1  
US-09-783-590-762/c  
Sequence 762, Application US/09783590  
Patent No. US20020110850A1  
GENERAL INFORMATION:  
APPLICANT: Dillion, Patrick J.  
APPLICANT: Haseltine, William A.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16.2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 762  
LENGTH: 445  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (44)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (52)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (62)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (106)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (123)

```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (345)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (389)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (423)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (426)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (441)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (445)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-390-762

```

```

Query Match
Best Local Similarity 41.9%; Score 26.8; DB 10; Length 445;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 9 AGTCGGCTGATAGGCTTGAAAGTTTCATTTCCAGACCCATTATTACATGCTAG 62
DB 232 AGTCCCGAGATTGGCTTGAGCAATTCACATTCATTAACCCCTTGATACAGGTAG 179

```

```

RESULT 2
US-09-918-995-30115/c
; Sequence 30115, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30115
; LENGTH: 362
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(362)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30115

```

```

Query Match
Best Local Similarity 38.4%; Score 24.6; DB 9; Length 362;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 2 AAAAAAAGTCGCTGATAGGCTTGAAAGTTTCATTTCCAGACCCATTATTACATGCTA 61
DB 343 AAAAGAAAAAGATGATAGATAGCTTTGATCTCATTTCTTCTTATTCATGTA 284
QY 62 GCC 64
DB 283 GCC 281

```

```

RESULT 3
US-09-918-995-7625/c
; Sequence 7625, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7625
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(394)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-7625

```

```

Query Match
Best Local Similarity 38.4%; Score 24.6; DB 9; Length 394;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 2 AAAAAAAGTCGCTGATAGGCTTGAAAGTTTCATTTCCAGACCCATTATTACATGCTA 61
DB 363 AAAAGAAAAAGATGATAGATAGCTTTGATCTCATTTCTTCTTATTCATGTA 304
QY 62 GCC 64
DB 303 GCC 301

```

```

RESULT 4
US-09-918-995-8119/c
; Sequence 8119, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8119
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(425)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8119

```

```

Query Match
Best Local Similarity 38.4%; Score 24.6; DB 9; Length 425;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 2 AAAAAAAGTCGCTGATAGGCTTGAAAGTTTCATTTCCAGACCCATTATTACATGCTA 61
DB 224 AAAAGAAAAAGATGATAGATAGCTTTGATCTCATTTCTTCTTATTCATGTA 165
QY 62 GCC 64
DB 164 GCC 162

```

```

RESULT 5
US-09-918-995-31162/c

```





APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SAME, AND METHODS OF USE  
CURRENT APPLICATION NUMBER: US/09/938,842A  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 3526  
LENGTH: 1245  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3526

Query Match 38.1%; Score 24.4; DB 9; Length 1245;  
Best Local Similarity 68.0%; Pred. No. 29;  
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 5 AAAAAGTCCGCTGATAGGCTGAAAAGTTCATTCAGACCCATTTT 54  
DB 1011 AAAAAGTCCGCTGATAGGCTGAAAAGTTCATTCAGACCCATTTT 1060

## RESULT 9

US-08-781-986A-551/c  
Sequence 551, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 551:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4923 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-551

Query Match 38.1%; Score 24.4; DB 7; Length 4923;  
Best Local Similarity 73.8%; Pred. No. 48;  
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 16 TGATAGGCTGAAAAGTTCATTCAGACCCATTTTACAT 57  
DB 4116 TGATAGGCTGAAAAGTTCATTCAGACCCATTCATTAATAT 4075

## RESULT 10

US-09-974-300-7252/c  
Sequence 7252, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085,500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7252  
LENGTH: 1043  
TYPE: DNA  
ORGANISM: Bacillus clausii  
US-09-974-300-7252

Query Match 37.8%; Score 24.2; DB 10; Length 1043;  
Best Local Similarity 66.0%; Pred. No. 33;  
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2 AAAAAGTCCGCTGATAGGCTGAAAAGTTCATTCAGACCCATTTT 54  
DB 159 AGAAAATCTCCAGTGAAGAGGCTGCAACCGCTTACTAGCCCATGATG 107

## RESULT 11

US-09-974-300-4354  
Sequence 4354, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085,500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4354  
LENGTH: 1228  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
US-09-974-300-4354

Query Match 37.8%; Score 24.2; DB 10; Length 1228;  
Best Local Similarity 66.0%; Pred. No. 35;  
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1 CAAAAGTCCGCTGATAGGCTGAAAAGTTCATTCAGACCCATTTT 53  
DB 176 CAAAAGTCCGCTGATAGGCTGAAAAGTTCATTCAGACCCATTTT 228

```
RESULT 12
US-09-878-574-796/c
; Sequence 796, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 796
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB028-046-Q1-B1-B10
US-09-878-574-796

Query Match
Best Local Similarity 37.2%; Score 23.8; DB 10; Length 380;
Matches 37; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2 AAAAAAGTCCGCTGATTAAGCTTGAAGTTCATTCCAGACCCATTTCATCGT 60
DB 183 AATAAAAAGCACACAAAGAAAGCTACTAATTTTCATTAGTGCATCTTTTACATGCT 125

RESULT 13
US-09-815-242-6976
; Sequence 6976, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6976
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
```

```
NAME/KEY: CDS
; LOCATION: (1)...(1371)
US-09-815-242-6976

Query Match
Best Local Similarity 37.2%; Score 23.8; DB 10; Length 1371;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 AAAAAAGTCCGCTGATTAAGCTTGAAGTTCATTCCAGACCCATTTCATTTT 52
DB 1051 AAAAAATCTACGCTGTGAAGCTTCTAAGTAATCACTGACCTATGTT 1101

RESULT 14
US-10-003-806-10/c
; Sequence 10, Application US/10003806
; Patent No. US2002011929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02060S1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Version 3.1
; SEQ ID NO 10
; LENGTH: 173808
; TYPE: DNA
; ORGANISM: Mouse
US-10-003-806-10

Query Match
Best Local Similarity 36.9%; Score 23.6; DB 12; Length 173808;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 CAAAAAAGTCCGCTGATTAAGCTTGAAGTTCATTCCAGACCCATTTCATTCGT 60
DB 116783 CAAAAAAGTCCGCTGATTAAGCTTGAAGTTCATTCCAGACCCATTTCATTCGT 116724

QY 61 AG 62
DB 116723 AG 116722

RESULT 15
US-09-896-852-45/c
; Sequence 45, Application US/09896852
; Patent No. US20020025542A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Maine, Gregory T.
; APPLICANT: Hunt, Jeffrey C.
; APPLICANT: Brojanac, Susan
; APPLICANT: Jhy-Tsing Sheu, Michael
; APPLICANT: Chovan, Linda E.
; APPLICANT: Tyner, Joan D.
; APPLICANT: Howard, Lawrence V.
; TITLE OF INVENTION: ANTIGEN COCKTAILS, P35, AND USES THEREOF
; FILE REFERENCE: 6361.05.D1
; CURRENT APPLICATION NUMBER: US/09/896,852
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/086,503
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 4451
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
```

Fri May 30 09:39:40 2003

us-09-990-099-23.rnpb

Page 6

US-09-896-852-45

Query Match 36.6%; Score 23.4; DB 10; Length 4451;  
Best Local Similarity 73.2%; Pred. No. 1.1e+02;  
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCGCTGATTAAGGCTTGAAAAAGTTCATTCCA 42  
|||||  
Db 1686 AAAAAAAAAAGTCCGCTGATTAAGGCTTGAAAAAGTTCATTCCA 1646

Search completed: May 5, 2003, 18:05:48  
Job time : 48.3626 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:53:33 ; Search time 390.505 Seconds  
(without alignments)  
2654.281 Million cell updates/sec

Title: US-09-990-099-23  
Perfect score: 64  
Sequence: 1 caaaaaaacgcgcgtgata.....cccatcttaccatcgtacc 64

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST\*\*  
1: em\_estba:\*  
2: em\_estlin:\*  
3: em\_estlin:\*  
4: em\_estlin:\*  
5: em\_estlin:\*  
6: em\_estlin:\*  
7: em\_estlin:\*  
8: em\_estlin:\*  
9: em\_estlin:\*  
10: em\_estlin:\*  
11: em\_estlin:\*  
12: em\_estlin:\*  
13: em\_estlin:\*  
14: em\_estlin:\*  
15: em\_estlin:\*  
16: em\_estlin:\*  
17: em\_estlin:\*  
18: em\_estlin:\*  
19: em\_estlin:\*  
20: em\_estlin:\*  
21: em\_estlin:\*  
22: em\_estlin:\*  
23: em\_estlin:\*  
24: em\_estlin:\*  
25: em\_estlin:\*  
26: em\_estlin:\*  
27: em\_estlin:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
C 1	38	59.4	863	17	AF094932 AF094932
C 2	28.4	44.4	700	9	AJ425333 AJ425333
C 3	28.4	44.4	750	9	AL564193 AL564193
C 4	28.2	44.1	379	10	AV820230 AV820230
C 5	28	43.8	215	10	BB014998 BB014998
C 6	27.8	43.4	518	10	AV699330 AV699330

C 7	27.8	43.4	580	13	BJ529330 BJ529330
C 8	27.8	43.4	620	10	AV662150 AV662150
C 9	26.8	41.9	384	14	TS1247 TS1247
C 10	26.8	41.9	680	9	AT061603 AT061603
C 11	26.6	41.9	740	17	AQ054300 AQ054300
C 12	26.6	41.6	742	17	AQ086715 AQ086715
C 13	26.6	41.6	934	12	BG443824 BG443824
C 14	26.4	41.2	288	9	AA343029 AA343029
C 15	26.4	41.2	335	14	T67877 T67877
C 16	26.4	41.2	355	14	T67504 T67504
C 17	26.4	41.2	559	13	BI775172 BI775172
C 18	26.4	41.2	559	13	BI775175 BI775175
C 19	26.4	41.2	569	9	AA046980 AA046980
C 20	26.4	41.2	587	13	BM288511 BM288511
C 21	26.4	41.2	1005	17	CNS0021A CNS0021A
C 22	26.2	40.9	290	14	TA1012 TA1012
C 23	26.2	40.9	307	13	BM158733 BM158733
C 24	26.2	40.9	379	14	NS8056 NS8056
C 25	26.2	40.9	407	10	AV649451 AV649451
C 26	26.2	40.9	440	9	AI285962 AI285962
C 27	26.2	40.9	477	9	AA723197 AA723197
C 28	26.2	40.9	501	9	AA723197 AA723197
C 29	26.2	40.9	504	9	AI065114 AI065114
C 30	26.2	40.9	553	12	BG565748 BG565748
C 31	26.2	40.9	606	12	BF241003 BF241003
C 32	26.2	40.9	633	10	AV699943 AV699943
C 33	26.2	40.9	639	10	AV682396 AV682396
C 34	26.2	40.9	755	12	BG564841 BG564841
C 35	26.2	40.9	282	14	T74865 T74865
C 36	26.2	40.6	415	14	H54054 H54054
C 37	26.2	40.6	435	12	BG044593 BG044593
C 38	26.2	40.6	604	9	AI541136 AI541136
C 39	26.2	40.6	656	17	BH11596 BH11596
C 40	26.2	40.6	673	17	AG173379 AG173379
C 41	26.2	40.6	1097	12	BE908674 BE908674
C 42	26.2	40.6	1108	12	BE908541 BE908541
C 43	26.2	40.6	1398	14	BQ952447 BQ952447
C 44	25.8	40.3	222	14	T57263 T57263
C 45	25.8	40.3	230	9	AA342380 AA342380

## ALIGNMENTS

RESULT 1  
LOCUS AF094932/c 863 bp DNA linear GSS 29-AUG-2000  
DEFINITION AF094932 Salmonella typhimurium LT2, Lambda DASH II Salmonella typhimurium genomic clone 1335-T7, DNA sequence.

ACCESSION AF094932.1 GI:4322774  
VERSION AF094932.1  
KEYWORDS GSS.  
SOURCE Salmonella typhimurium.  
ORGANISM Salmonella typhimurium  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 863)  
Wong R.M.Y. and McClelland M.  
End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,  
TITLE Li-Cor

JOURNAL  
COMMENT Unpublished (1999)  
Contact: McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3099 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@sfsc.sdsu.edu  
Class: Shotgun.

FEATURES  
source  
1. 863  
/organism="Salmonella typhimurium"  
/strain="LT2"  
/db\_xref="taxon:602"  
/clone="1335-T7"

RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel.: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@rct.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

us"

50 a 59 g 50 g 56 t

ORIGIN					
Query Match	43.8%	Score 28;	DB 10;	Length 215;	
Best local similarity	66.7%	Pred. No. 5.9;			
40: Conservative	0;	Mismatches	20;	Indels	0;
40: Conservative	0;	Mismatches	20;	Indels	0;
40: Conservative	0;	Mismatches	20;	Indels	0;

Db 90 CAATAAAAAGGCTGTGGTCTTGAATGCAAAAGTCTTCTCCAGACTCATGCTTGAATCC 31

REFERENCE  
YU X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.  
1 (bases 1 to 518)  
Eukaryota; Metazoa; Primates; Catarrhini; Hominoidea; Homo.  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 518)  
YU X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.  
1 (bases 1 to 518)  
Eukaryota; Metazoa; Primates; Catarrhini; Hominoidea; Homo.  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 518)

TITLE	Xiao, H., Gu, J., Liu, F., Han, Y., Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, A., Hu, G., Gu, J., Chen, Z., and Han, Z.
JOURNAL	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
COMMENT	21625106
	Contact: Zeguang Han
	Chinese National Human Genome Center at Shanghai
	351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
	201203, P. R. China
	Tel: 86-21-50801919(ex.45)

FEATURES  
SOURCE

1. 518  
Location/Qualifiers

Query Match  
Best Local Similarity 43.4%; Score 27.8; DB 10; Length 518;  
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Db 2 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTA 61  
81 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTA 140

QY 62 GCC 64  
11  
Db 141 GCC 143

BASE COUNT 148 a 97 c 122 g 150 t 1 others

ORIGIN

RESULT 7  
LOCUS BJ529330/c 580 bp mRNA linear EST 09-AUG-2002  
DEFINITION BJ529330 MF01SSB CDNA Oryzias latipes CDNA clone MF01SSB035105 5',  
ACCESSION BJ529330  
VERSION BJ529330.1 GI:22188142  
KEYWORDS EST  
SOURCE Japanese medaka.  
ORGANISM Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.  
REFERENCE Kohara, Y., Shin-I, F., Kimura, T., Narita, T., Jindo, T. and Takeda, H.  
TITLE Medaka EST Project in Takeda's lab  
JOURNAL Unpublished (2001)  
COMMENT Contact: Tadasu Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source

1. 580  
Location/Qualifiers

/organism="Oryzias latipes"  
/strain="Hd-r"  
/db\_xref="taxon:8090"  
/clone="MF01SSB035105"  
/clone\_lib="MF01SSB CDNA"  
/sex="mixture of female and male"  
/tissue\_type="whole embryo"  
/dev\_stage="segmentation stage 20 - 25"  
BASE COUNT 168 a 100 c 97 g 209 t 6 others

Query Match  
Best Local Similarity 43.4%; Score 27.8; DB 13; Length 580;  
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATA 56

Db 251 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATA 197

RESULT 8  
LOCUS AV662150/c 620 bp mRNA linear EST 16-JAN-2002  
DEFINITION AV662150 GLD Homo sapiens CDNA clone GIDAA02 3', mRNA sequence.  
ACCESSION AV662150  
VERSION AV662150.1 GI:9883164  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,  
Xiao, H., Gu, J., Lu, F., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,  
Hu, G., Gu, J., Chen, Z., and Han, Z.  
TITLE Insight into hepatocellular carcinogenesis at transcriptome level  
JOURNAL by comparing gene expression profiles of hepatocellular carcinoma  
MEDLINE with those of corresponding noncancerous liver  
COMMENT Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex. 45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn

FEATURES  
source

1. 620  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GIDAA02"  
/clone\_lib="GLD"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 227 a 103 c 114 g 176 t

ORIGIN

Query Match  
Best Local Similarity 43.4%; Score 27.8; DB 10; Length 620;  
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTA 61  
363 AAAAAAAAAAGTCGATGAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTA 304

Db 62 GCC 64  
11  
Db 303 GCC 301

RESULT 9  
LOCUS T51247 384 bp mRNA linear EST 06-FEB-1995  
DEFINITION y03e12.s1 stragene placenta (#937225) Homo sapiens CDNA clone  
IMAGE:70126 3', similar to gb:A06977 SERUM ALBUMIN  
PRECUSOR (HUMAN). mRNA sequence.  
ACCESSION T51247.1 GI:653107  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



```

/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB0066M06"
/clone_1ib="CGUI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI.
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a

```

haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu)).

BASE COUNT 224 a 157 c 123 g 235 t 1 others  
ORIGIN

Query Match 41.9%; Score 26.8; DB 17; Length 740;  
Best Local Similarity 68.5%; Pred. No. 21;  
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Oy 1 CAAAAAAGTCGCTGATAGCTTGAAGTTCATTTCCAGACCCATTTTA 54  
Db 639 CAAAGAAATTCCTTTATACGGATCTAGGATTAATTTCCAAACCATTTCTTA 692

RESULT 12  
LOCUS A0867715 742 bp DNA linear GSS 03-NOV-1999  
DEFINITION nbe0034K11r CUGI Rice BAC Library (ECORI) *Oryza sativa* genomic clone nbe0034K11r, DNA sequence.  
ACCESSION A0867715  
VERSION A0867715.1 GI:6218276  
KEYWORDS GSS.  
SOURCE *Oryza sativa*.

ORGANISM *Oryza sativa*.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; *Oryza*.  
1 (bases 1 to 742)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)  
Seq primer: GGAAACAGCTATGACCAG  
Class: BAC ends  
High quality sequence start: 23  
High quality sequence stop: 460.  
Location/Qualifiers

FEATURES  
source  
1..742  
Location/Qualifiers

/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbe0034K11r"  
/clone\_lib="CUGI Rice BAC Library (ECORI)"  
/issue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBACindigo; site\_1: EcoRI; site\_2: EcoRI; rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three

times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu)).

BASE COUNT 249 a 137 c 134 g 221 t 1 others  
ORIGIN

Query Match 41.6%; Score 26.6; DB 17; Length 742;  
Best Local Similarity 70.0%; Pred. No. 25;  
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 2 AAAAAAAGTCGCTGATAGCTTGAAGTTCATTTCCAGACCCATT 51  
Db 210 ACANNTAGTCTCTGATAGCTGTGCAAACTGAACCTTCACACACTTT 259

RESULT 13  
LOCUS BG443824 934 bp mRNA linear EST 15-MAR-2001  
DEFINITION GA\_Ea0022E01f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum CDNA clone GA\_Ea0022E01f, mRNA sequence;  
ACCESSION BG443824  
VERSION BG443824.1 GI:13353476  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Gossypium arboreum.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 934)  
Wing, R.A., Fritsch, D., Yu, Y., Mao, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
An integrated analysis of the genetics, development, and evolution of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)  
Seq primer: TAAATCGACTCACTATAGCG  
High quality sequence stop: 665.  
Location/Qualifiers

FEATURES  
source  
1..934  
Location/Qualifiers

/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0022E01f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/issue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="Vector: pBK-CMV; site\_1: EcoRI; site\_2: XhoI"

BASE COUNT 276 a 212 c 233 g 206 t 1 others  
ORIGIN

Query Match 41.6%; Score 26.6; DB 12; Length 934;  
Best Local Similarity 78.0%; Pred. No. 26;  
Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 23 GCCTGAAAGTTCATTTCCAGACCCATTTTACATCTGTAC 63

DB 383 GCTTGAGAGGTCATTCAGACCATTTTTCACCGTAGC 423

# RESULT 14

AA343029

## LOCUS

DEFINITION

AA343029 388 bp mRNA linear EST 21-APR-1997  
EST48667 Fetal spleen Homo sapiens cDNA 3' end similar to similar  
to albumin, mRNA sequence.

## ACCESSION

AA343029

## VERSION

AA343029.1 GI:1995265

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 288)

## AUTHORS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulton, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., McInerney, J.O., Gocke, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.-C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glock, A., Kelley, J.C., Liu, L.-I., Marmore, S.M., Merrick, J.M., Moren, Palanques, R.F., McDonald, L.A., Nguyen, D.T., Pelligri, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sander, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Dime, D., Feng, D.F., Ferris, A., Fischer, C., Coleman, T.A., Collins, E.J., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.D., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

## TITLE

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

## JOURNAL

Nature 377 (6547 Suppl.), 3-174 (1995)

## COMMENT

96026280

## FEATURES

source

1. 288

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):144514"

/db\_xref="taxon:9606"

/clone\_lib="Fetal spleen"

/dev\_stage="Fetus"

/note="Organ: spleen; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT

85 a 53 c 52 g 91 t 7 others

## ORIGIN

Query Match

Best Local Similarity 41.2%; Score 26.4; DB 9; Length 288;

Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

DB 102

62 GCC 64

DB 162

GCC 164

# RESULT 15

T67877

## LOCUS

DEFINITION

T67877 335 bp mRNA linear EST 22-FEB-1995  
IMAGE:83067 3' similar to gb:A06977 SERUM ALBUMIN PRECURSOR (HUMAN  
); mRNA sequence.

## ACCESSION

T67877

## VERSION

T67877.1 GI:679025

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 335)

## AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chissole, S., Dietrich, N., Dubuque, T., Favella, A., Gish, W., Hawkins, M., Hollman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

## TITLE

Generation and analysis of 280 000 human expressed sequence tags

## JOURNAL

Genome Res. 6 (9), 807-828 (1996)

## COMMENT

97044478

## FEATURES

source

1. 335

/organism="Homo sapiens"

/db\_xref="GDB:500124"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:83067"

/clone\_lib="Stratagene liver (#937224)"

/sex="male"

/dev\_stage="49 years Old"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: liver; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Hepatotomy from normal male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT

84 a 73 c 63 g 99 t 16 others

## ORIGIN

Query Match

Best Local Similarity 41.2%; Score 26.4; DB 14; Length 335;

Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

DB 29

62 GCC 64

DB 89

GCC 91

Search completed: May 5, 2003, 18:41:09  
Job time : 395.505 secs

